全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

HZ BOOKS 华章经管

WILEY

统计套利

股指期货・股票联动・对冲・投资组合・长期获利 必读

STATISTICAL ARBITRAGE

ALGORITHMIC TRADING INSIGHTS AND TECHNIQUES

（美）安德鲁・波尔（Andrew Pole）著
陈雄兵 张海珊 译



机械工业出版社
China Machine Press

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

统计套利

STATISTICAL ARBITRAGE

ALGORITHMIC TRADING INSIGHTS AND TECHNIQUES

（美）安德鲁·波尔（Andrew Pole）著
陈雄兵 张海珊 译



机械工业出版社
China Machine Press

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

推荐序一



2010年对于中国金融市场来说，具有里程碑意义。3月31日，沪深交易所融资融券交易试点正式启动；4月16日，首批四个沪深300股票指数期货合约正式上市交易。融资融券及股指期货的正式推出意味着中国金融市场有了真正的做空机制，为金融市场的发展带来了许多有利条件和新的历史机遇，也提出了严峻的挑战。抓住机遇、迎接挑战，需要尽快研究做空机制下的交易策略，统计套利正是这方面经典的应用。一方面，统计套利策略的实施依赖于存在做空机制的资本市场；另一方面，投资者参与股指期货和融资融券也需要统计套利这样的多元操作策略为其提供指导。

统计套利是运用数量化的方法构建投资组合，在降低市场波动率的同时，获得投资回报。统计套利从业人员从以往的股价变动模式中发现规律和趋势，然后利用这些规律，在市场中寻找微小的价格差异，从事套利交易。打个比方，应用统计套利的方法研究工商银行和建设银行股票价格波动规律，发现这两家公司的股票价差大于1元的情况很少。当市场的价格波动导致建设银行的股票价格比工商银行高出1元，可以通过卖出建设银行股票和买进工商银行股票构建投资组合，当两家公司的股票价差回归到1元以内时做相反的交易，从而获得交易收益。无论是建设银行被高估还是工商银行被低估，都可以获得交易利润。适度的统计套利交易对于整个金融市场而言，有助于降低市场的波动性，提高资源的配置效率，减少市场系统风险。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

《统计套利》一书描述了统计套利的发展历程，运用生动的案例介绍了一些重要理论和算法模型。本书主要内容可以概括为以下五点：首先，追溯了统计套利策略的起源——匹配交易（pairs trading）的基本原理，阐释了其主要特性，匹配交易也成为贯穿全文的一个重要概念；其次，在承认模型有效性的前提下，论述了一些重要的时间序列模型，从最基本的加权移动平均模型，到复杂的动态因子分析模型；第三，专章论述了许多重要的理论概念，如爆米花理论、反转理论、突变理论等；第四，回顾了统计套利在经历了 15 年的辉煌以后陷入困境的原因；第五，理性地分析了统计套利复兴的背景和原因。

《统计套利》一书的作者安德鲁·波尔先生具有多年运作统计套利对冲基金的管理经验，本书作为其研究成果，是理论与实践相结合的典范。本书并不像一些畅销书那样通俗易懂，而是充满了复杂的公式和数字，但通过阅读本书，你可以探究统计套利的真正含义；了解统计套利的发展历程。更重要的是，知晓统计套利的运作方式和获利机理后，敏锐的投资者可以借此在金融市场中捕捉获利的机会。

理论物理学家用数学公式阐释宇宙运行的规律。牛顿三大定律完美地描述了这个世界的机械运动原理，麦克斯韦方程准确刻画了光波和电磁波的传播规律。统计套利从业者采用物理学的方法和数学的公式来描述现实金融市场。但能将金融市场当成一部精密的机器吗？证券的价格是由数学公式决定的吗？这些数学上的公理和定理应用到金融市场中，能产生有用的结果吗？记得爱因斯坦说过："我并不假装了解宇宙——它比我大得多（I don't pretend to understand the universe——it's much bigger than I am）。"我们也不能假装完全了解金融市场，要深入理解统计套利模型的运作机理，谨慎地使用统计套利模型。

金融行业面对的市场是全球统一的，我们需要学习和借鉴西方的做法和经验。《统计套利》一书吸收了金融理论和实践的最新成果，采纳在实践中出现的最新案例和统计资料，体系完善，论据充分。我相信，由陈雄兵和张海珊博士翻译的这本书，对丰富我们的交易理念，提高金融交易管理水平，将起到积极的推动作用。

潘功胜

中国农业银行股份有限公司执行董事、副行长

2010 年 9 月 7 日

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net



推荐序二

正如大多数人类活动一样，价格回归均值的反转现象是一种强大的力量，它驱动着系统与市场形成自我平衡。从20世纪80年代早期开始，在追求利润行为的驱动下，人们产生了将行为模型化的愿望，统计套利在此时成为一种正式而成功的尝试。要想理解复杂的金融工程和风险模型带给金融领域的发展和进步，那么理解统计套利（有时简写为 stat. arb.）的算法机理就成为一个很重要的基石。

统计套利所涉及的交易策略，一般都当成一种难以理解的投资规律。这种观点认为，它被两种互为补充的力量驱使，这两种力量来源于规律的核心特征：从业人员的困惑和部分投资者缺乏数量知识。统计套利从证券价格的系统性波动中，运用数学模型产生利润。没有一个投资经理愿意泄露复杂的交易方法。投资经理在挑选股票时，会讲一个很好的故事，但不会透露他们的决策方法。这与数量金融学家以模型为基础的策略不同。任何一个有意义的细节都会迅速应用到一系列的实验中，一个思维敏捷的听众都能试着对这个策略进行逆向工程。数量金融的从业人员只谈具备数学知识才能理解的概念，这也就是其中的原因。

对于想要寻找资产管理人员的投资者来说，对于数学知识熟练程度的要求，增加了统计套利的不透明性。要想理解一个统计套利从业人员说的是什么，不仅仅是字面的意思，人们需要具有超出大学水平的高等数学知识。很显然，这

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

限制了很多的对象。由于缺乏可供学习的参考资料，这个限制更加难以克服。现在，《统计套利》填补了这一空白。

统计套利已经存在了25年。在这期间，早期的从业人员用讲故事的方式，将一般的概念，传播给对此感兴趣的投资银行分析师和学术界人士。由于从业人员在不断提升模型的复杂程度，而且由于商业的原因，他们的创新也不对外透露，从而导致外界并不了解统计套利。在广泛传播的基本概念中，均值回归及它的变形体和均值反转等概念非常引人注意。均值反转是一个简单的概念：身材非常高的父母所生的孩子，一般会比他们的父母矮；而身材非常矮的父母所生的孩子，一般会比他们的父母高。对于大多数人来说，这是个很容易理解的概念。将这个观点应用到证券价格的波动中，意味着证券价格会返回到平均值。迄今为止，一切都运转正常。但是，我们遇到一个问题，身高的反转是两代人之间的生理现象，而价格反转是一个实体的动态过程。

价格会从哪里开始反转呢？均值是什么？成年人的平均身高是一个很熟悉的概念，如果想要精确地量化，也需要做一些工作。小学生可以根据他们所知道的成年人的平均身高，通过演绎的方式，合理地估计成年人的平均身高。在证券价格中，并没有可供利用的一般观测值或经验值。证券的涨跌幅度是任意的，价格反复波动。而且它们可能会崩盘，价格变为零。人类不会长得像天一样高，然后又反转到平均值，但证券价格却可能出现这种现象。

即使我们假设这个问题已经有了合理的答案，但又出现了其他专业性的问题：如何识别价格偏离了平均值呢？偏离多少？回归均值需要多长的时间呢？

现在，讨论中存在不透明性，并且很难清除它。数学模型的语言与一些不熟悉的概念混合，产生一种不安的感觉，是一种由于缺乏理解的恐惧感。

在《统计套利》中，波尔为读者提供了一个关于统计套利基本原理的教学旅行，消除了统计套利的不透明性。在20世纪80年代与90年代早期，对于多数人来说，都不了解统计套利。十多年之后的今天，统计套利的世界更加广泛，也更加复杂。

这与自然世界一样，经过了40亿年的进化之后，形成了复杂得令人难以置信的社会组织。然而，最简单的社会组织仍然很繁荣，而且到目前为止，拥有这个星球上最大数量的生物。在统计套利中，同样存在这种情况，一些基础的方法仍然支撑着统计套利的实践。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

《统计套利》描述了一些现象、产生这些现象的驱动力量、利用机会的动态发展模式，以及将均值反转运用到证券价格中的模型。它还提供了一些好的建议，从建立更精密的模型的线索，到对建立模型与监视绩效的忠告，这些忠告不仅仅适用于统计套利。

第1章和第2章描述了统计套利的起源，包括20世纪80年代年代令人崇敬的匹配交易以及这些机会具有巨大的扩展性和生产力。这些知识拉开了统计套利理论发展的序幕，提出在实践中应用这些交易规则的方法以及校验方法。在第5章中，论述了更多关于不透明性的内容，说明了逐日观察到的证券价格反转与每日价格的统计分布两者之间的关系。

第8章和第9章讲述了统计套利的中年危机。从2000年安然公司崩盘开始，到2001年恐怖分子攻击，以及波尔所说的公司“令人震惊的”不当行为，美国的金融市场动荡了好几个月，对于统计套利的绩效产生了巨大的影响。此外，在金融市场中产生了一些技术上的变化，包括采用十进制方式报价，以及在纽约证券交易所中独立的交易人员减少等原因。波尔说明了为什么统计套利的绩效会中断。这种中断还没有结束，给人深刻的印象。

第10章和第11章讲述了统计套利的未来。交易算法一开始破坏了经典的统计套利方法。但现在，波尔认为交易算法是一种在系统上利用交易机会的开端。他将一种新的运动命名为“突变过程”：用突变理论来说明动态过程的模型，突变理论是行为模式一个合理的解释。统计套利复兴，给人留下深刻的印象。

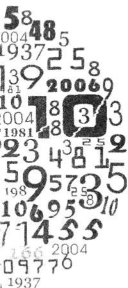
《统计套利》的风格非常直接和彻底，不存在任何困惑。有时候，它也会采用自由的风格，比如在第11章中用蝌蚪命名的小故事。散文式的风格也随处可见。

在描述数学模型时，作者们很容易使用不容易记住的、公式化的措词，除了代数能提供的解释或说明之外，什么也不能提供。《统计套利》是个例外，它打破了“不透明”的乌云，这乌云也是波尔要回避的均值吧！

格雷戈里·万·基普尼斯

2007年4月23日于纽约

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net



前　言

本书讲述的是统计套利的故事。它既描述了统计套利的历史，追溯始于20世纪80年代统计套利策略在摩根士丹利诞生，到面临重重考验的21世纪初的过往岁月，同时也阐释了统计套利的运作方式和存在的原因。这些介绍来源于一些基本的原理，并在最大程度上采用一个基本的分析框架。当阐述技术原理时，我们努力抵制各种诱惑，希望为广大感兴趣的读者，呈现一个易于理解、容易接受的成果。我只能是“尽可能”，因为第7章和第11章的附录属于“未能抵制诱惑”。对于许多顶尖专业人员构建的大多数模型，本书只是对其概念进行了论述，而且举例也仅仅局限于读者所熟悉的模型上。作为统计套利的起源，匹配交易（pair trade）这一术语在本书中更多地是被作为一个有教育意义的例子，而不是承认它在实际交易中发挥的作用。正因为采用的是这种方式，所以读者有可能忽视了本书作为一部匹配交易手册的作用，虽然读者阅读这部作品时自身具有的经验、阅读的目的以及对本书的期待具有广泛性，但发生上述情况也是不可避免的。在实际交易中，简单且缺乏精心设计的匹配交易不再有利可图，但是，匹配交易在解释说明性方面依然是有价值的工具，因为它在避免将问题变得复杂的前提下，保留了阐明观点、构建模型和分析问题的能力。在市场历经了25年的发展之后，读者要构建一个超越本书理解和说明范围之外的有利可图的交易策略，需要在结构的复杂性和分析的精巧性上多做文章。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

本书采用的精巧设计方案是构造了一组相似的匹配交易，这些交易通常被指定为一个群体，但是，对于如何运用矩阵来测量其相似程度的细节，并没有过多地探究。构建这样的股票组合模型可以采取几种方法。一些人更愿意将所构建的股票组合定位在通用模型上，然后观察实际匹配交易与通用模型的偏差，进而调整预期；另外一些人则喜欢采用正式的数学公式，将其作为一种合成工具。近20年来，这两种方式以及其他一些方式都可以被作为评级模型进行正式的分析，这在主流统计学思想领域非常盛行，而且在见解和应用方面都卓有成效。将时间序列中的动态因素加入到标准静态结构中，能非常便捷地构造一个精巧的分析结构，其复杂程度可能超过本书经常提起的工具。尽管如此，所有这类模型的发展都依赖本书所详细阐述的观点和技巧。

具有高等数学和统计学知识的读者，能迅速找到应用知识的场合。

本书的重点内容集中在结构简单的匹配交易上，希望读者将本书看做是一本手册，能清晰地找到“如何做”的方法。从这个角度来讲，读者从本书中能了解到统计套利发展的历史，包括“什么是统计套利，它是如何运作的以及为什么它具有合理性”。同时，也正如我们前面所提到的那样，要成功地执行交易还要求读者进行更多的思考和应用性研究。要完成这一任务，你可以将本书当做一张地图，它揭示了统计套利的主要特性，并且暗示读者在哪里需要使用指南针和笔记本。当你进行此类冒险时，要记住地图绘制人员标注的“危险”这样的标记，因为这对你可能会非常有用。

本文坦然地把统计学家的观点（即模型具备有效性）当做是正确的。坚持模型的有效性是本书的主题之一。统计学家对变量的理解，即认可这样的看法：尽管有用，但其模型是错误的，而且这种错误的性质是通过结构性“误差”（观测值和模型预测值之间的差异）来阐明的，这也是本书的另外一个主题。可以说，这不仅仅是一个简单的主题，更是统领全书、具有导向性的核心思想。

第1章介绍了匹配交易的概念，第2章对匹配交易进行详细说明。通过解释说明和举例，针对匹配交易和反转现象，我们提出两个简单的理论模型。这些模型将贯彻全文，被用来研究可能性是什么，说明怎么利用这些可能性，考虑什么样的变化会导致负面影响，并描绘这种影响的特性。本书同时描述了如何选择一系列金融工具进行建模和交易的方法。在开始分析的时候，就需要考虑“变化”这个因素，因为动态时间变化会对整体投资方案起到强化作用。没

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

有动态变化，就不存在套利。

在第3章中，增加了分析的深度和广度，将建模的范围从简单的观测性的匹配规则㊀，扩展到正式的统计模型，适用于更具普遍意义的投资组合。本书描述了几个流行的时间序列模型，但讨论的重点集中在两个模型上，一种是极端复杂的加权移动平均模型，另外一种是因素分析模型，尽可能将这两种极端情况下的模型所具备的信息论述清楚。正如前面已经提到的，匹配价差模型将贯穿全书，作为最简单的实际应用案例用于概念讨论。在有必要特别强调的地方，我们会直接提出套利者所关心的各方面的问题，包括投资组合的优化方法和风险因子的暴露程度。但是，在大多数情况下，本书尽量避免涉猎多个领域的知识。波动率模型（以及极富吸引力的随机共振模型）将在第3章和第6章中专门予以讨论，在本书的其他章节将纳入到均值预测过程的探讨中。

第4章提出概率定理，这是一种描述股票价格运动的流行方法，这种方法在20世纪80年代晚期首次应用到简单规则之后，就经常被拿来使用。了解这种结果可以指导模型开发策略中的价值评估工作。因为这一结果受结构性的动态变化驱动，通过长期在公共领域认真观察就可予以揭示，是否需要建模人员的智慧结晶，或者只是像高中毕业生会做的那样，只是简单执行就可以了？许多统计套利从业人员声称将会获得高比例收益的模型。他们所建立的模型与基本价差模型或（重要的）风险管理模型比起来都显得过于简单，需要在精巧性上花更多的功夫。当市场崩盘并且违背理论成立的假设条件时，比较统计套利从业人员的实际绩效，我们可以发现他们对统计套利的



注　释

㊀ 这种说法并没有轻视的意思：简单的观测性的匹配规则是非常有效的。统计的范围被限制在衡量变量的变化范围上；先努力得出可能的观测值，然后再进行分布研究、建立模型公式、进行估计、分析误差，或者进行预测。在收集大量观测值后，就可以开始上述研究工作。随着统计研究的扩展，将交易经验加入到对历史数据的分析中，对股票价格波动的微妙之处会有更深入的认识，市场力量也会驱动套利机会反复出现。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

基本知识的理解，存在着显著的差异。当事情一切顺利时，无论是具有理论知识的管理者还是盲目操作的人，所能获得的回报都是相同的，但当理论成立的假设条件不成立时，理论知识通常能揭示更多的信息（托尼·欧·哈根认为概率定理的基本结论是众所周知的，但我已无法验证这种说法的正确性。这个结论是如此微不足道，以至于没有一个正式的名称，它仅仅作为一个简单的结论，或者作为练习题出现在概率分布的教材中。无论如何，对于统计套利来说，这个结论具有非常深远的意义）。

第 5 章针对一篇已出版的文章展开评论，以验证反转现象发生的非限制性条件（文中我们不提作者的名字，以避免引起尴尬）。文章忽略了正态分布的核心作用，提出两个孪生的错误主张：①股票价格序列一定要显示出正态边际分布的特征，才可能发生反转现象；②一个正态边际分布的序列，必然会发生反转的现象。对于这两种观点，给予了驳斥。如同第 4 章所阐述的一样，反转现象无处不在。

第 6 章对“在一个价差组合中，波动率究竟是多少”这个问题予以了解答，因为在进行反转交易时，它直接关系到量化可利用的机会。

第 7 章讲述的是存在诸多象形文字的概率微积分，非常适合热衷该内容的爱好者。只要是曾经认真上过概率理论课程的人，都应该能够理解该章所提到的观点，并且大多数人也应该都能看懂其中的推论。概率微积分的机理并非十分复杂。刚开始的时候，我们对一些概念进行了区分，对读者来说可能具有一定的挑战性——建议多读两遍！案例中蕴含着重要而富有实际意义的见解，努力领悟一定会获得回报。至于理论上的抽象概念是如何反映在实际价格序列的测量特性上，这一知识对于评价建模的可能性、模型仿真或者交易结果来说都是非常珍贵的。不过，尽管第 7 章非常重要，但要提醒读者的是，阅读本书其他部分不需要本章的知识。尽管你可能会漏掉随后讨论中一些更精巧的观点，但并不会因为没有理解第 7 章的内容而看不懂其他部分的论述。

从第 8 章到第 10 章，我们或许可以给其贴上“衰落”的标签，因为这几章所讨论的内容具有共同的特性，即描述了从 2000 年开始统计套利遇到

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

问题，并且直接引起了2002～2004年收益的灾难性下降。我们从历史中吸取的深刻教训是，在2000年的时候，并没有某个单一的条件或者是某几个条件组合发生剧烈的变化，使统计套利模型没能达到预期的绩效。那么究竟是怎么回事呢？它比现实更具有戏剧性，可以说是诸多复杂的原因和时机相互融合之后产生的结果。许多为人所熟悉的观点，包括报价单位采用十进制、竞争、低波动率等因素都试图解释这一切。每个说法都有一定的影响力，但是其中任何一个单独拿出来，或者甚至是将这些因素综合起来，都无法影响金融市场。市场的价格动态变化确实发生了根本性的改变，不管是频繁发生的每日交易，还是超过一个月以上的投资都受到了影响，这大大地削弱了反转机制的经济价值，对于这种情况，我们需要进行更深入的研究。

一个变化接着一个变化就是第9章的主要内容，也是2002～2004年统计套利无法获得收益的根本原因（在2000～2002年期间，业绩恶化非常明显，但仅限于行业内一部分从业人员。但是到了2002～2004年，统计套利策略普遍无法获取收益）。直到最近，美国宏观经济历史中很不寻常的插曲终于结束了，但是它所造成的后果仍然影响着美国金融市场，数以百万计的投资者的集体行为是随机的；当然，不管以何种方式存在于市场，投资者继续在体验着那些变化，以及引起变化的原因。

交易方式发生了转变，从纽约证券交易所热闹的大厅，逐渐变成无声的电子交易，通过大型经纪商及投资银行设计的计算机算法，慢慢积聚力量，最终变成类似冰河时期一样无法平息的巨变。这些变化缓慢、巨大、不能抗拒，颇具毁灭性，同时也带有推陈出新的特性[⊖]。在动态变化中，市场波动



注　释

⊖ 根据信用市场中技术的发展以及交易型开放式指数基金（ETF）的发展，得出一个主要结构性结论。联邦储备银行（FED）实际上也重新检查了以前的价格行为模式，将计算机算法的交互作用作为股权交易的动因纳入考虑范围。除了前述考虑之外，随着证券交易委员会条例与纽约证券交易所规则的变化，改革现状也同时列上了日程。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

率的下降频繁地被拿来讨论。市场波动率到哪里去了？可以说，有很大一部分是被算法给“吞噬”了。个体参与者的声音被弱化了，不过，混乱的场面事实上并没有受到影响。有相当多的市场参与者和混乱的场面，以另外一种方式继续存在着。就目前的情况而言，计算机程序（见第10章）“管理”着超过60%的美国股权交易，其所产生出来的效果，就有点像是在给市场这个多动症患儿打了一剂“利他林”（ritalin）。有两种关于低波动率的论述：其一是，悲观地认为市场缺乏凯恩斯所说的动物性本能，主要的说法是，当亚洲正在复兴时，美国的创业激情却受到了限制；其二是，投资者忘记了投资决策的风险，以及与此相伴而生的恐惧，低估了波动率，做出错误的决策，在将来产生负面影响。这两种说法不一致，但这种矛盾可以被化解。与第一种说法相对立的观点是，那种动物性本能依然很活跃，每天约有15亿股的股票在纽约证券交易所进行交易。对此，我们还能得出其他的结论吗？或许应该说，这种动物性的本能仍广泛存在。算法是没有情感的。因此，出现了大量创新性的风险承担方式，按照历史标准进行比较，其波动率相对偏低，这是由交易技术引起的，但尚未被许多市场参与者认识到。如果以历史数据来检验现有的波动率水平，不可避免地高估了风险。

暂不考虑统计套利机会的进化与第10章之间存在什么关系，就第10章本身而言还是很有趣的。算法和计算机驱动的交易正以多种多样的方式改变着金融世界。世界上绝大多数人都已经选择电子交易方式进行交易——谁知道再过几年，纽约证券交易所的大厅会不会变成博物馆、停车场，或者只是出现在我们的记忆中呢？

第11章描述了由于算法交易的技术进步，导致统计套利浴火重生，重新创造一个新的局面。新型的股价动态模式已经浮出水面。统计套利的故事又回到了新的起点，这只雏鸟能否顺利地飞翔呢？

在第11章中，预测统计套利的复兴，出现在2005年，如今预测已经在逐渐变成现实。至少从2006年早期开始，那些经历了2003~2005年富有挑战性的动态变化过程的从业人员，已经看到了其业绩表现。有趣的是，虽然

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

在普通股票的价格运动中，存在新的系统性模式，但旧的模式似乎也已经恢复了活力。人们正在加速使用算法交易，目前有20家以上的公司能够提供相应的工具软件。在技术进步的发展过程中，从2006年下半年开始，至少有两只以算法为基础的对冲基金出现在市场中。对于统计套利从业人员而言，这是一个令人兴奋且恰逢其时的好时机！

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

Andrew Pole. Statistical Arbitrage: Algorithmic Trading Insights and Techniques.

Copyright © 2007 by Andrew Pole.

This translation published under license. Simplified Chinese Translation Copyright © 2011 by China Machine Press.

No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system, without permission, in writing, from the publisher.

All rights reserved.

本书中文简体字版由 John Wiley & Sons 公司授权机械工业出版社在全球独家出版发行。未经出版者书面许可，不得以任何方式抄袭、复制或节录本书中的任何部分。

本书封底贴有 John Wiley & Sons 公司防伪标签，无标签者不得销售。

封底无防伪标均为盗版
版权所有，侵权必究
本书法律顾问 北京市展达律师事务所

本书版权登记号：图字：01-2010-3263

图书在版编目（CIP）数据

统计套利/(美) 波尔 (Pole, A.) 著；陈雄兵，张海珊译. —北京：机械工业出版社，2011.1

(金融知识堂)

书名原文：Statistical Arbitrage: Algorithmic Trading Insights and Techniques

ISBN 978-7-111-32544-4

Ⅰ. 统… Ⅱ. ①波… ②陈… ③张… Ⅲ. 数理经济学－应用－金融学 Ⅳ. F830

中国版本图书馆 CIP 数据核字（2010）第 224386 号

机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码 100037）
责任编辑：刘 斌　　版式设计：刘永青
北京瑞德印刷有限公司印刷
2011 年 1 月第 1 版第 1 次印刷
170mm × 242mm · 14.5 印张
标准书号：ISBN 978-7-111-32544-4
定价：38.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：(010) 88379210；88361066
购书热线：(010) 68326294；88379649；68995259
投稿热线：(010) 88379007
读者信箱：hzjg@hzbook.com

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net



目 录

推荐序一

推荐序二

前言

第 1 章 蒙特卡罗的谬误 …… 1

1.1 起源 …… 1

1.2 未来的方向 …… 4

第 2 章 统计套利 …… 8

2.1 导论 …… 8

2.2 噪声模型 …… 9

2.3 爆米花过程 …… 17

2.4 识别匹配交易 …… 19

2.5 投资组合结构和风险控制 …… 24

2.6 动态变化和校验 …… 30

第 3 章 结构模型 …… 35

3.1 导论 …… 35

3.2 正式的预测函数 …… 37

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

3.3 指数加权移动平均模型 …… 38
3.4 古典的时间序列模型 …… 45
3.5 哪一类回报？ …… 51
3.6 因子模型 …… 52
3.7 随机共振 …… 58
3.8 实践中的事情 …… 59
3.9 加倍交易：更深入的探讨 …… 62
3.10 因子分析入门 …… 64

第4章 反转定律 …… 68

4.1 导论 …… 68
4.2 模型和结论 …… 69
4.3 非齐次方差 …… 75
4.4 一阶序列相关性 …… 77
4.5 非常数分布 …… 82
4.6 结论的应用 …… 84
4.7 应用于美国债券期货 …… 84
4.8 总结 …… 86
附录4A 向前预测几天 …… 87

第5章 高斯不是反转之神 …… 89

5.1 导论 …… 89
5.2 双峰骆驼与单峰骆驼 …… 90
5.3 依然在敲响钟声 …… 94

第6章 价差波动率 …… 97

6.1 导论 …… 97

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

6.2 理论上的解释 …… 101

第7章 将反转机会量化 …… 110

7.1 导论 …… 110
7.2 平稳随机过程中的反转现象 …… 111
7.3 非平稳过程：不均匀的方差 …… 131
7.4 序列的相关性 …… 132
附录7A 在示例6中对数分布的一些细节 …… 134

第8章 诺贝尔的困惑 …… 135

8.1 导论 …… 135
8.2 事件风险 …… 136
8.3 一个新的风险因素的出现 …… 139
8.4 赎回压力 …… 142
8.5 《公平披露条例》 …… 144
8.6 在亏损期间的相关性 …… 145

第9章 多重困难 …… 148

9.1 导论 …… 148
9.2 十进制 …… 149
9.3 统计套利结束了 …… 152
9.4 竞争 …… 153
9.5 机构投资人 …… 155
9.6 波动率是关键因素 …… 155
9.7 关于时间维度的思考 …… 158
9.8 虚构情节中的真实示例 …… 165
9.9 恶劣的行为 …… 165

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

9.10 对2003年的剖析 …… 168
9.11 结构变化的真实情况 …… 169
9.12 总结 …… 170

第10章 黑匣子出现 …… 172

10.1 导论 …… 172
10.2 对交易成交量期望值和市场冲击力进行的模型化 …… 174
10.3 动态更新 …… 177
10.4 更多的黑匣子 …… 178
10.5 市场紧缩 …… 178

第11章 统计套利的复兴 …… 180

11.1 突变过程 …… 183
11.2 突变预测 …… 187
11.3 趋势变化的识别 …… 189
11.4 突变过程在理论上的解释 …… 194
11.5 风险管理的含义 …… 199
11.6 结束 …… 201
附录11A 理解Cuscore统计量 …… 201

致谢 …… 211
译者后记 …… 212
参考文献 …… 213

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net



第 1 章 蒙特卡罗的谬误

对于重复出现的事件，无论它们看起来多么奇怪，我们必须时刻准备着去学习和研究。

——著名统计学家　博克斯

1.1　起源

1985 年，在天体物理学家努齐奥·塔塔里亚（Nunzio Tartaglia）㊀的指导下，一小群擅长定量研究的科研人员创造了一个用匹配组合方式买卖股票的程序。摩根士丹利的“黑匣子”由此诞生，在迅速赢得声誉的同时也赚得盆满钵满。统计套利，这个在当时十分生僻的词汇，由此开始了 15 年的辉煌历程。


注　释

㊀ 在《威尔莫特论文精选》（*The Best of Wilmott*）一书中，保罗·威尔莫特指出，摩根士丹利的匹配交易程序是在 1982 ~ 1983 年由杰瑞·班贝格创造出来的。班贝格在 1985 年离开摩根士丹利，到普林斯顿纽波特合伙企业工作，于 1987 年退休。我们不能确定班贝格在摩根士丹利的程序与塔塔里亚的差异。另外还有一些人认为这是一个团队的功劳，仅将团队的带头人称为“发明者”是不公平的。

有趣的是，威尔莫特声称早在 1980 年他的公司就发明了匹配交易。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

"黑匣子"的细节对外是保密的，但是臆测流传于坊间，其基本规则很快昭示天下，"匹配交易"一词也开始出现在金融词典上。匹配交易的假设前提非常简单，即寻找一对有相同历史价格行为的股票，当两只股票的价格存在较大偏离时，断定这一价差随后必会趋于收敛。这个规则简单、明了，而且拥有巨大的获利能力。

你一定想知道塔塔里亚的见解源自何处。正如同多数发明故事一样，需求是激发创新的动力。在大宗交易中，管理团队在进行获利交易的同时，需要找到一种对冲风险的方法。塔塔里亚运用自身的数学素养提出一个主张，即可以通过卖空具有相似交易行为的股票，来规避大宗交易风险。这种方法一经问世，便被越来越普遍地应用到匹配交易的创新中。很快，一个新的利润空间由此形成。

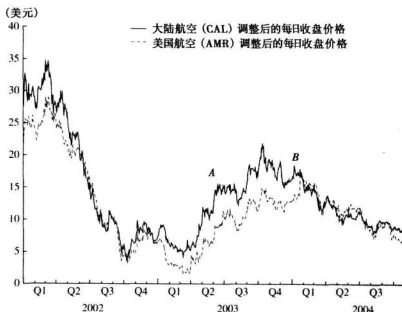


图 1-1 大陆航空和美国航空（2002～2004 年）的每日收盘价格

从图 1-1 中，我们可以看到大陆航空和美国航空自 2002～2004 年每个交易日的收盘价。注意这两只股票的运行轨迹，其价差时而较大，时而较小。匹配交易方案所倡导的就是：当价差较大时（如 *A* 点），买进价格较低

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

的股票，同时卖出价格较高的股票；当价差较小时（如 B 点），则做相反的交易，清空头寸。

1985 年的时候，计算机还没有像现在这样普及，每日股票价格只能依靠专门的工具进行接收。匹配交易要求纯粹的数字运算能力，其实现的关键是配备价值数万美元的硬件设施。匹配交易在概念上简洁宜用，在多年后的今天也很容易付诸实践，但在其诞生初期，只有投资公司才有可能真正地去研究和运用它。

那个时期的许多故事在业界流传，也赋予相关的企业和从业人员传奇色彩。其中有两个故事是真实的，至今仍具有重要意义。一个是美国证券交易委员会（SEC）至今还在使用这一算法来甄别异常的价格行为模式；另外一个则是原本持反对意见的专家，实现了从最初的质疑到完全接受这一方法的彻底转变。

同其他人一样，美国证券交易委员会也被围绕着摩根士丹利“黑匣子”的光环所吸引。在了解模型是如何预测特定股票价格时，人们很快意识到，这项技术能被用于识别异常的或非法的股票价格波动。很久以后，神经网络技术才扮演了这一角色。

在 20 世纪 80 年代晚期，纽约证券交易所（NYSE）拥有的独立专业交易商超过 50 个。当摩根士丹利开始系统地进行“买弱”和“卖强”交易时，大量资本有限的家族企业则对此持有强烈的怀疑态度。他们最大的担心是摩根士丹利这一大机构会操纵其他小的专业机构。随着匹配交易模式的真相大白，这些怀疑由安心取而代之，最后则是完全的接受。以至于当专业交易商看见摩根士丹利开始建仓一只“走势较弱”的股票时，就会立即趋之若鹜，因为他们达成了“共识”：这只股票今后一定会上涨。

早期，这项策略带来了巨大收益，它迅速地催生出了一些成功的独立交易商，包括 D. E. Shaw 和 Double Alpha，这两家著名的基金公司都是由塔塔里亚以前的助手创立的。在接下来几年，其他团体创造了大量匹配交易业务，而其创始人要追溯到摩根士丹利，例如戴维·肖的公司。随着匹配交易

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

的普及，学院派的兴趣被激发出来，通过美国国家经济研究局（NBER）以及其他机构出版的文章，这个基本概念为更多人所知道。同时个人计算机因成本降低而大量增加，使得潜在参与者的基数呈现爆炸式增长。很快，实际参与者的人数也确实实现了爆炸式的增长。

1.2 未来的方向

20年以后，从“婴儿期”的匹配交易成长、成熟起来的统计套利正面临着剧烈的环境变革。收益回报已经大大减少。统计套利管理者的境遇可谓困难重重，需要调整策略来应付现状。进入新世纪，金融市场环境的变化使管理者面临着生死存亡的挑战，就如同冰河时期降临时，远古生物所面临的环境一样。行动敏捷且适应能力强的生物得以存活下来，行动缓慢或者转型慢的生物，则被冻僵或饿死。

统计套利的冰河时期始于2000年，并在2004年进入“冰点”。评论员称这一投资规律将就此消失，投资者撤回资金，机构倒闭，全面的溃退席卷而来，失败的阴影笼罩在统计套利的上空。

但我相信，判断统计套利将就此终结还为时尚早。尽管由于市场结构变化，传统的统计套利模型遇到了一些问题（这些问题将会在接下来的章节逐一描述），但是新的机会已经显露端倪。至少在两个高频率时间维度，出现了新的股票价格行为模式。在电子交易商的相互影响下，这种驱动力量敲响了美国股票交易未来的钟声。

不得不承认，现在仅能对出现的新机会进行大致的特征描述，但其具有非常重要的经济价值，并可能足以使统计套利避免灭亡的命运。这就好像古代克罗马努人终将被现代人所取代一样。本书将在第11章对新的机会进行展望和阐述。

我曾经考虑将本书的题目命名为《统计套利的兴起、衰落和复兴》，以反映统计套利的历史和现在可能出现的机会。在本书接下来的章节中，这种模式清晰可见。本书几乎是以编年史的方式编写。对于那些急于知道“统计

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

套利的前途在哪里"的读者来说，从第 1 章到第 7 章一直在讨论统计套利历史和理论发展，似乎有些不合时宜，并且无需关注。这就好比建议一个研究应用数学的学生去研究哥白尼的天体运动学说，并告诉他那是具有实用价值的一样。但我坚持认为研究历史是有价值的（对于数学家也一样，能够从中进行很多类似的比较）。了解统计套利之前做了什么，以及它是如何运作、为什么这么运作，对于理解市场结构的变化对统计套利产生的负面影响提供了必要的基础。了解哪一种变化起了作用，并且那些作用是如何实现的，能让我们明白在当前情况下，还可以预见什么。

借助过去的情况解释现在的现象并不是一个新的概念，它是科学调查的合理基础。很多人都很熟悉政治哲学家的警告：不研究历史的人，将注定重复犯同样的错误。㊀但那不是我们的关注点。毫无疑问，一些从事套利的人曾经犯过一些个人的错误，但不能由此就断定所有从业人员都有问题，并对套利理论加以全盘否定。我们的关注点是更加引人注目的科学观点："站在巨人的肩膀上。"不考虑价值判断、科学理论对错，也无论贡献多么微小，都应该对过去的理论进行详细的研究。了解市场变化是如何导致原本有利可图的套利变得毫无价值，能够让我们理解并评价新出现的机会是否具有前途。

让我们阐述得更清晰一些。尽管有科学天才曾经提及，但历史上没有专门的文献支持本书描述的工作。与物理学、化学和数学不同，不容易证明统



注 释

㊀"人类的进步不在于不断的改变，而是依赖于对过去的记忆。当变化不可逆转之时，一切无需改进，也没有留下可能的改进方向：经验无法保留下来，就如同身处野蛮人之中，如同永远无法度过婴儿期。不能记住过去的人，注定会重复过去的错误。在生命的最初阶段，心智是轻佻且容易分散的，无法保持连续性和持续性，错过了进步的机会。这是小孩和野蛮人的现状，在他们的直觉中无法从经验中学习知识。"出自《理性的生活》(*The Life of Reason*)，乔治·桑塔亚纳（George Santayana）。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

计套利领域的研究工作的正确性。它与经济学、社会科学更接近，因为它们基本的驱动力量都是人。我们可以给股票价格上突然出现的波动贴上“反转”的标签，然后用时间模式进行描述，再将这些模式用数学公式简洁地表达出来，但并以此指导我们的行动（也就是交易）。但是，理论、模型和分析方法仅仅是对随机出现的过程进行描述，并没有对因果关系给出正确的论述。不管我们如何描述从业人员采用的流程和程序，从分析师（写报告）到基金顾问（读报告、推荐投资组合变化），到基金经理（进行投资决策），直到交易员（按决策进行交易），建模都是基本的流程。在统计套利的起源阶段，在探索过程中的模型化结论都难逃失败的命运。令人吃惊的是，这些失败的经验为统计套利复兴的种子提供了养料。

不像历史学或政治哲学的研究，一定浸透着个人的见解，这些见解会随着新文物的发现或者对权威文献的质疑而发生变化，统计套利研究的数据是不可更改的、明确的、完整的历史数据，任何学者都可以获得。证券价格的历史数据是不可改变的，就像布拉赫（Brahe）的天文观测值。尽管布拉赫的记录受到时间以及现实物理世界中必然存在的不确定性的限制，但证券的历史价格却是精确已知的。㊀

我们在检验证券价格的数据质量时，要注意的一点是，布拉赫是用引证和演绎出科学理论，检验对宇宙规则的影响。我们记录金融交易的数据可能毫无错误，但它们是对人与人之间交易博弈结果的衡量。不变的物理规律应用到金融交易数据中，会发生什么？我们可以建立价格变化的模型，但是当


注 释

㊀ 约翰·弗兰斯蒂德（John Flamsteed，1646—1719）是第一个皇家天文学家，他在格林尼治建立的皇家观测台系统地记录了天空的变化，编制了30 000份单独的观测报告，每一份报告记录并见证了在40年里那些努力工作的夜晚。在达娃·索贝尔（Dava Sobel）的《经度》（*Longitude*）一书中提到：“丹麦天文站所编制的星系记录是泰戈·布拉赫所编制记录的三倍，这在很大程度上提高了统计数据的精确性。”

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

我们这么做时，数据基础可能发生变化。数据无法更改，但同时也不会重复出现。在这种情况下，我们如何做才能科学地验证一个理论是有效的?

这些问题在此无法给予解答。没有人能从哲学或社会学的角度来解释金融学。但是，我们可以力求在数据分析、前提假设、模型建立和检验方面做到科学严谨。这种严谨性是一个人提出信念主张的基础，它支持我们去正确理解金融领域某个组成部分突然出现的特性，并采取合理一致的行动。

本书对什么是统计套利进行了批判性分析。统计套利是一种针对概率和数量的正式理论，对反映在金融市场中的美国经济结构的巨大转变给予了解释，并将视角专注于套利可能产生的惊人后果上。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

第 2 章 统计套利

许多实际发生的状况，都可以被轻易地视为随机变化，但是，有时隐藏在变化中的重要信号可能是要警告我们，出了什么问题，或是提示我们出现了新的机会。

——博克斯

2.1 导论

以塔塔里亚团队所从事的研究为开端，匹配交易方案的制订也开始朝着不同的方向发展。因为研究者所使用的分析技术越来越复杂，所开发的模型越来越偏重高深的技术，所以这种越来越为人们所熟悉的技术交易规则，开始被冠以深奥的称号。“统计套利”一词被首次使用是在 20 世纪 90 年代早期。

统计套利方法的范围，从最古老的纯粹的匹配交易机制到复杂的、动态的非线性模型，应用的技术包括神经网络（neural networks）、小波分析（wavelets）、分形分析（fractals）——几乎涵盖了统计学、物理学和数学上所有的模式匹配技术，这些技术被测试、检验，并在大多数情况下遭到摒弃。

统计套利后期的发展融合了多种因素，包括交易经验、更多的实证观察

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

值、实验分析，并且从工程学和物理学的视角（涉猎领域从高能粒子物理到流体动力学，应用的数学技巧从概率论到微分差分方程），给予了理论上的解释。随着越来越多的智力成果被应用到相关的研究中，“匹配交易”这一名称似乎不太恰当了，甚至有些过时。“统计套利”一词由此被创造出来，它更能引起人们的好奇心，尽管并没有统计学家参与，甚至大部分工作都未涉及统计的内容。

2.2 噪声模型

匹配交易的第一个预测规则，是用简单的数学表达式来描述价差的图形外观。如图2-1所示，大陆航空和美国航空的股票价差落在 -2 美元到 6 美元之间，其中一个简单而有效的规则是，当价差为 4 美元时，建立一个空头头寸，当价差是 0 美元时，则对冲平仓。

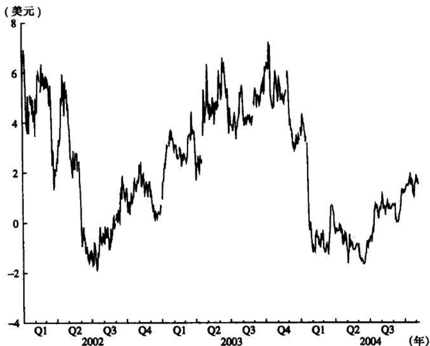


图2-1 大陆航空和美国航空每日收盘价差

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

我们特意使用“规则”这一名称，而非模型，原因在于仅仅是对所观测的行为的主要特征进行描述，并没有试图构造流程去解释这一行为。这样说并不是要贬低这项规则的有效性，而是对统计套利早期的工作特征给予更精确的阐释。正如历史记录所显示的那样，这一规则曾在几年中为使用者带来了令人难以置信的丰厚利润。

将“4 美元—0 美元”这个规则应用到大陆航空和美国航空的价差交易中，在 2002～2003 年期间，就会出现一次匹配交易的机会。这是不费吹灰之力的赚钱方式，这也是塔塔里亚在 1985 年发现的商机。更令人吃惊的是，数千种股票都呈现出匹配交易的特征。

当人们看到价差出现这种情况，得出第一个简单的规则后，试图去找出其他的规则。另外两个交易规则是：

规则 2　对反向的情况进行交易。

规则 3　在更强的进场时点，重复进行交易。

2.2.1 反向交易

在 2002 年下半年，当价差从很小的数值向 4 美元这个进场点逼近时，为什么我们不进行交易？为什么不对这样的价差变化采取行动呢？采用“海龟交易法则”（turtle trade）的商品交易商，很快用规则 2 取代了规则 1。将“价差在 0 美元时清仓”这一离场条件进行修改，变成反向交易，“在价差为 0 美元时，建立相反的多头与空头头寸”。现在，交易者会一直持有头寸，等待价差从一个较低的位置上涨，或从一个较高的位置下跌。

对交易机会进行拓展，在不增加额外工作量的情况下，带来了更多的交易机会和更大的利润。

2.2.2 多重交易

在 2002 年的第 1 季度，大陆航空和美国航空的股票价差从 7 美元掉到 1 美元，有 6 美元的变动范围。如果按照规则 1（与规则 2）来交易，将经历市场波动带来的重大收益和损失，但是由于仅有一次进场交易信号，因此并不能获得任何收益。既然在这几天或几周内，价差会上下波动，或增加或减

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

少，而到最后价差变为零，出现了退出信号（规则1），或者反转信号（规则2），那么为什么不试着从这些变动中捕捉一些套利的机会呢?

在规则1的基础上设计出来的规则3，增加了两个进场时点，目的就是从价差中获得更多的收益。就大陆航空和美国航空而言，当价差增加到6美元时，建立第二个头寸，认为价差会逐渐缩小。这样，2002年和2003年对价差缩小的交易数量会成倍增加，最后将会增加150%的利润（在2002年，规则2产生的利润只占很小的比重，那是因为规则3并没有影响，规则2的反向交易所获得的利润。在2003年没有反向交易，这些头寸也就进入了2004年）。

这个简单的例子，用非常清晰的方式描述了在1985年呈现在塔塔里亚团队面前的巨大机会。在那个时代，价差变化的幅度通常比本章所展示的例子要大得多。

2.2.3 规则的校验

当有人想对这个简单的匹配交易进一步分析，或者对这个交易进行更长时间的检验，就会涉及校验问题。图2-2中展示了另外一个匹配交易在2000年的交易历史，图2-3展示了相对应的价差。⊖



注 释

⊖ 由于美国航空在2000年3月16日将预订机票的业务分离出去，成立了萨博(Sabre)公司，因此公司的股票价格序列应据此进行调整。如果不进行调整，股票的收盘价格在一个晚上，将从60美元直接掉到30美元，出现如此剧烈的变化。我们按时间顺序向后调整价格，保留最近的价格，使分离前的价格与实际市场中的价格不同。在2000年1月对美国航空股票进行交易，应当按分离之前的际价格60美元进行交易。向前或向后调整股票价格，尽管调整的方式必须一致，完全取决于个人偏好。根据调整后的历史价格，计算出来的投资回报的时间序列与调整的方式无关，是唯一的。大多数分析都是用“回报”的方式来完成的，而不是采用价格。由于用价格表述观点更容易理解，在本书中采用价格作为示例。对于公司所发生的事件，包括股息和拆分，在计算交易获利时，需要对股票价格进行调整。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

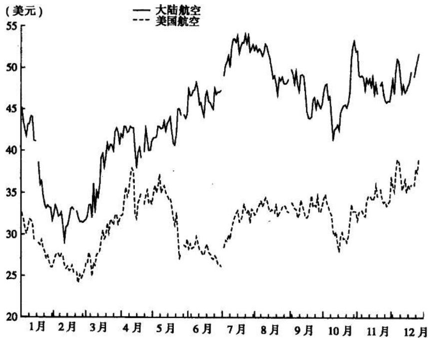


图 2-2 大陆航空和美国航空的每日收盘价格（2000 年）

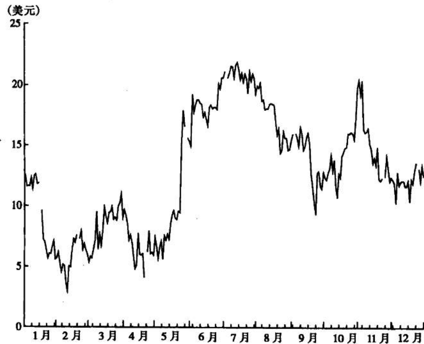


图 2-3 大陆航空和美国航空的每日收盘价差（2000 年）

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

我们早就应该研究这个例子。图 2-3 显示大陆航空和美国航空的例子中，价差的变化超过 20 美元，存在 3 次交易机会。如果应用前三个规则，第一个困难是：前面的结论在这个示例中是不起作用的。如果应用规则 3，在这个示例中将存在两个交易点，在 1 月当价差超过 4 美元和 6 美元时。在 7 月时，当价差增加到 20 美元时，对之前的头寸产生很大的压力（账面亏损严重）。直到这一年结束，在账面上依然存在损失。很清楚，我们必须为规则 1 至 3 制定不同的校验方式。同样清楚的是，这些规则的基本形式依然是有效的。

现在，需要考虑的是校验的方法，应用到数百或数千条潜在的价差交易中。用眼睛来对图像进行判断，那么需要很多双眼睛。如果采用数字化的程序，需要一个自动化的校验规则。我们采用统计学的方法。通过观察价差的变化范围，决定交易规则。对自动计算的方式而言，这是一件很容易的事情：在图 2-1 中，价差的最大或最小值分别是 -2 美元和 7 美元，设定 20% 的边界，根据规则 1 自动计算出进场的价格为 5 美元与 0 美元。这与我们用眼睛观察的数值不完全相同，自动计算方式会产生相似的交易结果（尽管复杂一些）。对计算机而言，不管价差是多少，这个程序都可以重复使用。

以第二个图（图 2-2）为例，价差的范围是 3～22 美元。以 20% 作为边界进行规则校验，进场和出场的价格分别为 18 美元和 7 美元。在 2000 年，将这个自动校验的方法应用到规则 1，能建立一个可以获利的交易。这比简单地采用之前的规则（通过观察的方法，从图 2-1 中得出在 4 美元与 6 美元时进场，在 0 美元时平仓），对图 2-2 中的价差进行交易导致损失来说，要好得多。

校验阶段：在上述采用目测的方式进行校验的讨论过程中，我们没有考虑时间跨度，在第一个例子中设定了两年的时间，在第二个例子中则设定了一年的时间。选择的这两个例子，是为了描述匹配交易机会的实用性。尽管如此，这两个例子还是切合实际的。我们要考虑的是，在匹配交易中，设定多长的时间跨度是适当的？

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

在图 2-1 和 2-2 中的股票是相同的，都是大陆航空和美国航空。“多长的时间跨度是适合的?”这个问题现在看来是要随着时间不断地进行拷问而不是仅仅在每次进行匹配交易时，做一次决定就可以了。想象一下，根据 2000 年的大陆航空和美国航空的价差进行校验，再使用所得到的结果用到 2002 ~ 2003 年的交易中。在这个例子中，结果看起来还是不错的，只是没有任何交易机会。但事实上，这是一个负面的结果，因为我们错过了一些有价值的交易机会。在其他的例子中，根据过时的历史数据对规则进行校验会造成很可怕的损失，付出高昂的代价。

使用多长时间的历史价格对交易规则进行校验，这是一个很关键的问题。目前采用的静态分析方法都是从历史价格中对规则进行校验，并应用到相同的历史数据之中。与之相比，在实际交易中，是将过去历史数据应用到未来之中。在图 2-4 中，揭示了大陆航空和美国航空过去四年（2000 ~ 2003 年）价差的历史数据，以及价差的上、下限。上、下限分别是采用回溯 3 个月历史数据中最大值和最小值的 20% 来设定的。尽管这些上下限与之前目测观察出来的上下限没有关联，但是它们为交易的识别保留了很好的方式。此外，与之前的计算示例不同，现在上下限的估值超出了示例所使用的方法。在任意一天，计算中用到的价格信息都是公开的而且可获得的历史数据。因此，这样计算出的价格信息是实际可行的。

如果应用规则 2，会产生 19 笔交易（没有考虑 2000 年第一季度的交易机会，因为没有充足的数据来计算上下限），其中有 4 笔亏损，15 笔盈利。不管最终是盈利还是亏损，这些交易在持有期间，对市场价格而言，都曾处于亏损状况。值得注意的一点是：从 2000 ~ 2003 年，价差的波动呈现下降的趋势，在以后的章节中将会重点论述。

2.2.4 为交易规则保留的价差预留界限

为了说明计算价差范围的上下限，是如何指导交易决策的，我们选择使用 20% 的预留界限。在 3 个月的价格中，“做空价差”是最大价差减去 20%，而“做多价差”是最小价差加上 20%。解释这个操作程序的含义

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

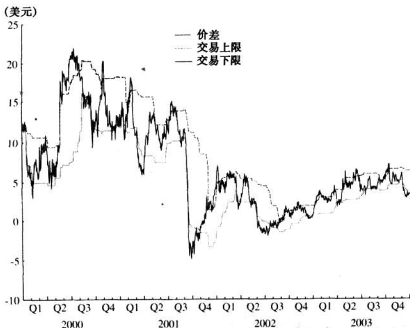


图 2-4 大陆航空和美国航空的每日收盘价差

上是非常便利的。它清楚地定义了交易预留的界限，也就是前 3 个月价差的 1/5。

令人不满意的是，这个交易规则应用了极值，也就是最大值和最小值。极值具有很大的可变性。采用极值会有很大的不确定性：回忆你读过的统计教科书知识，都告诉你要小心地使用极值。极值建模是一个复杂的研究领域，它的应用范围从预测洪水的最高水位，到预测电力需求和断电的可能性，以及其他各个方面。

由于极值的巨大的变动性，在实际交易规则中，需保留 20% 的预留界限。假设只选择价差范围的 10% 作为预留界限，希望在历史数据中产生足够多的交易机会。由于用过去的信息预测将来存在不确定性，因此在实际交易中，用 10% 的预留界限是不明智的。未来出现的极值，肯定与过去的极值不同。假如价差交易在“平静时期”，虽然有大量的获利机会，但识别出很少的机会，交易量会很少。比较好的方式是保守一些，使用一个较大的预

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

留界限。当然，在价差剧烈波动的时期，收入会随市场变化产生很大波动，但整个交易的利润不会下降。

采用极值以外的数值可以得到较好的稳定性。采用价值的均值，比采用极值，会获得更大的信心。因此，大多数应用，修正“做空”和“做多”的上下限，计算出相对均值的偏离度，而不是极值的偏离度。一个典型的例子是，布林线采用均值加减一个标准差作为上下限。尽管从理论上说，经过这样的转换，改善了多少稳定性是值得商榷的：计算标准差包含了所有的数据，也包括极值，由于所有的数据会取平方，极值实际上占很大的权重。在计算诸如均值和标准差这样的统计值之前，会采用一些增加稳定性的方法，比如去掉样本中的极值（在更精确的应用中，会采用降低权重的方法）。

根据价差分布将上下限往外推几个百分比，比如 20% 和 80%，也会产生相似的稳定效果，但是很少有人采用。从操作上讲，在这里描述的简单交易规则中，这种做法没有实践意义的。不过在一些更复杂的模型和不对称分布中，这种做法会有重要的意义。

在丧失一些可解释性的基础上，可获得较大的稳定性（这是一种假定）。分布的标准差和分布的范围之间并没有一定的联系。当应用标准差的方式时，即使没有意识到，其实是做了很多假设条件的，假定价差服从正态分布，均值加减一个标准差会有 2/3 的概率，均值加减两个标准差会有 95% 的概率。金融数据是非正态分布的，常常会出现非对称的情况，在距均值几个标准差的位置会出现较大的观察值。这就是所谓的“厚尾”现象。厚尾是与正态分布相比较而言的，而不是这些金融数据有什么特殊性情况。采用正态分布，计算尾巴代表的概率通常会错估风险，通常意味着低估。使用经验分布，也就是数据的分布，而不是假设的数学公式，大多数这种错误是能避免的。而且，检验正态分布的曲线，与一组数据的相似程度，并且判断计算概率的准确性，是很简单的，只要区分分布的中心或者尾部就可以。第 5 章将在讨论反转时，会利用时间序列来说明这些问题。

既然使用样本动差（均值和标准差），产生如此多潜在的错误，为什么

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

我们不使用价差范围值呢？在这样诡辩中，能获得什么好处？进行上述很多（几乎是无意识的）推论外，还有很多其他有意识的推论，这些推论是由数学模型所驱动的。极值（以及其函数）很难进行分析，而标准差容易得多。对于正态分布的假设而言，均值和标准差是分布的特征，在本质上是不可或缺的。

尽管从学术上来说，对于技术的理解和分析都是很重要的，但是20世纪80年代晚期和90年代早期的实际应用中，对技术的理解和分析仅占很小的比重：反转现象非常明显，幅度也很大，并且大规模地出现在股票中，以至于只有故意做错才会产生亏损。那样好的环境已经不存在很多年了。随着一些产业波动率的下降，公共事业产业是充分的例证（如 Gatev 等），粗糙的标准差规则不再是好的交易方式，因为期望的交易回报率出现萎缩，并低于交易成本。我们可以用设置最小的回报率下限的方法来解决这个问题，几年以后，这种做法也提供了一种有价值的风险管理工具。

2.3 爆米花过程

目前为止，所涉及的交易规则都强烈地认为，价差会在均值上下系统性地波动。这种随时间波动的模式，其原型是正弦波（sine wave）。在早期的匹配交易中，这种原型为价差分析提供了理论模型，但是也丧失了许多交易机会。如图2-5所示，一种被我们称为"爆米花过程"的可替代的原型，提供了一种新的见解。随着一个突然干扰的出现，价差会突然偏离均值，然后慢慢向均值回归。在这个模型中，取消了价差模型过度波动的限制（实际价差波动比数学模型更不规则）。一个往上的动作（向很远的顶点移动），可能在返回均值后，然后又向另一个顶点移动。另外一个类似的情况是，一个向下的运动随后跟着一个向下的运动，它们之间没有向上的运动。使用"很远的"限定词，是为了与偏离均值较小的波动进行区分。从定义上说，两个低点之间应该有一个高点来间隔。但是这个高点只有偏离均值足够大，并且能产生足够的经济价值，才是有价值的。一个介于两个低点之间的高点可能

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

接近均值，实质上并一定必须高于均值。

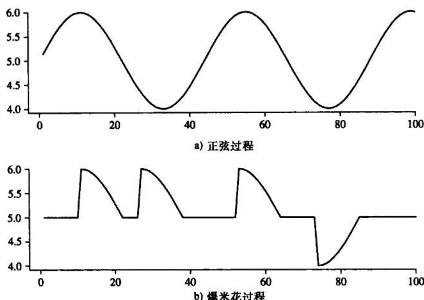


图 2-5 过程原型

用数学公式表示爆米花过程，会比写正弦函数要复杂，但是也不会复杂很多。例如正弦函数会写成：

$$y_t = \sin(t)$$

那么爆米花函数可以写成：

$$y_t = I_t \sin(t)$$

这里 I_t 是一个指示函数，不是 1 就是 -1，用于表示峰顶或者峰谷。在这里，数学不是最重要的；从对这个过程的描述和图形中，我们可以得出一个观点：如果使用海龟交易规则，在爆米花过程中应用是无效的。图 2-5b 中，海龟交易规则只能识别一笔利润为 2 美元的交易。爆米花过程提出了一个新的交易规则：当价差回归均值，就产生离场信号，而不是假设价差会从建立头寸的部位，超过均值向相反方向移动。这个新的交易规则识别出 4 笔交易，总利润为 4 美元。这个规则的新特征是，资金会有一段时期是闲置的。

这个新规则的所有必须的计算过程都描述得很清楚：需要的参数就是价

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

差的局部均值和价差的范围。下面是针对前述的状况对交易规则进行的调整。

规则4 当价差比均值偏差增加（减少）足够大（如 k 个标准差），卖出（买进）这个价差；当价差回到均值时，清空头寸。

从事统计套利的专业人员建立了很多更精巧的模型，不管是针对匹配交易的价差，或者是更复杂的股票历史价格所得到的函数，都是基于对爆米花过程的理解，或者是根据均值回归原理，而不是根据正弦过程或海龟交易过程。第3章中将描述一些这样的模型以及基于模型的延伸思考。随机共振现象（stochastic resonance，也在第3章描述）也是对规则4的前提条件进行了修正。

2.4 识别匹配交易

匹配交易蕴藏着巨大的机会。我们有了一系列可操作的交易规则和自动校验的程序。现在，我们可以识别哪些匹配交易能被执行了。

早期，股票按行业进行分类，每个分类中的成对的股票都是候选的交易机会。风险管理基本采用 Barra 模型，构建投资组合，然后从差异性很大的匹配交易投资组合中识别出因素暴露的偏差（例如使用股票或者标准普尔期货来抵消 β 风险）。

为了控制收益的波动性，引进了很多精巧的方法。这些方法也成为很多人向往的目标，同时，根据实际经验，也会显示出结构化方法的弱点。当对冲基金开始推广匹配交易和统计套利策略时，基金经理个人的喜好变得更有影响力。

选择相关性最大的成对股票，是一种最早用来选择匹配交易的方式：计算每一对候选交易（例如使用两年的每日交易数据）的相关性，然后留下相关性高于某一个最小值的匹配交易。在假设过去的相关性能用于预测将来的相关性的前提条件下，剔除了那些几乎没有相关性或相关性很小的匹配交易。该理论认为，没有相关性的股票在行为上也没有相关性，因此作为匹配

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

交易来说是不可行的。

2.4.1 精选匹配交易

当一项匹配交易涉及的两只股票的价格时而分开、时而聚合时，匹配交易关于反转的交易能够取得最好的效果。当股票 B 下跌时，股票 A 上涨，或股票 B 上涨时，股票 A 下跌，总之这种行为模式会产生非常低（甚至是负数）的相关性。从利润或者回报的角度来说，上面提到的相关性筛选方法（也就是寻找高度相关性的股票）不就是完全错了吗？事实不是这样的：短期看，剔除低相关性的匹配交易，可能会错过一些获利机会；但是从长期看，极大地改善了风险状况。与具有相同运动趋势的股票相比，那些出现明显相反或者不相关的价格运动的股票，在将来更有可能与市场整体的发展方向背道而驰。在某个时候，不相关的股票非常有可能产生代价高昂的匹配交易。

这种见解也产生了一种很精细的相关性筛选方法。当股票价格沿着不同的方向运动，形成一个峰顶或一个峰谷时，就将其定义为一次风险时刻（或事件）。以风险最小化为目标来选择匹配交易，可取的做法是，选择具有相似历史轨迹的股票，即这两只股票出现峰顶和峰谷的时间比较接近，而且在出现这些事件时，移动的大小程度也很相似。因为在市场遇到某种干扰时（例如政治上、产业发展上的干扰等），这些匹配的股票不大可能出现完全迥异的反应（可能需要剔除一些事件的余波）。而以利润最大化为目标来选择匹配交易的话，理想的做法是选择两只在事件与事件之间沿着不同价格轨迹发展的股票，这两只股票呈现出较强的负相关性——先是背道而驰，后又重新聚拢。关于匹配交易的相关性，什么样的符合需求，什么样的不符合需求，第 5 章将会给予正式的阐述。

2.4.2 事件分析

我们将股票价格出现转折点的情况当成发生了一次事件。识别转折点的算法方式如下：

（1）如果价格序列下降一个量，出现一个负回报，这个负回报的绝对值超过局部年度回报波动率的一个特定的比例，那么可以将局部的最大值当成

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

转折点。

（2）类似的情况，如果价格序列上升一个量，出现一个正回报，这个回报的绝对值超过局部年度回报波动率的一个特定的比例，那么可以将局部的最小值当成转折点。

（3）观察图2-6中的价格轨迹（通用汽车调整后的每日价格）。如果将 a 点的位置当成一个转折点，那么下一个转折点在哪里？a 点是一个局部最小值；因此下一个转折点必然是局部最大值。按时间顺序，观察从 a 到 t 之间的时间序列。识别 $[a, t]$ 之间的局部最大价格，用 p 点表示。从 p 点到 t 点的价格下降的程度，与 t 点（向后看）的局部波动率的百分之 k 相比，是否大一些？

（4）当 $p=m$ 和 $t=t_1$ 时，这一点不是转折点。当 b 点被识别为局部最大值（$t>t_2$），并且直到 $t=t_3$，b 点才被正式认定为转折点。

（5）在这个例子中，所用的参数是以20天为一个周期来定义局部波动率，年度回报因子为16，转折点的判断比例是30%。

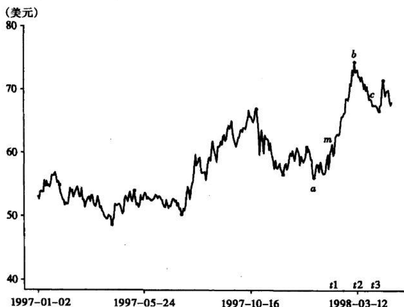


图2-6 调整后收盘价格轨迹（通用汽车），以30%作为转折点的识别标准

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

图 2-7 同样显示了通用汽车的价格序列，这次转折点的识别降低了标准：价格从峰顶下降的程度，只要达到局部波动率的 25% 就可以认定为转折点。与 30% 的标准对比可以发现，另外 4 个局部极值点被识别出来（不考虑序列最后的终点）。不过，在 1997 年年中的时候，有两个局部峰顶和峰谷没有被这个算法识别出来。它们提供的回报在几天内就达到了 4%，如果转换为年度回报，将是很惊人的。



图 2-7 调整后收盘价格走势（通用汽车），以 25% 作为转折点的识别标准

图 2-8 再次显示通用汽车的价格序列，只是采用了一个更低的标准：价格从峰顶下降的程度，只要达到局部波动率的 20% 就可以认定为转折点。如果跟 30% 的标准比较的话，另外 8 个局部极值点被识别出来（不考虑序列最后的终点），其中 4 个点，与 25% 的条件下识别的极值点一致。

在其他的例子中，改变周期的长度，同样根据一个降低后的标准，会识别更多的转折点。如果基于一个更高的标准，有些转折点就无法识别出来。这些例子和观察值提醒我们，这里的分析是严格的统计结果。这些事件反映了市场情绪，它们可能受到与股票不相关的新闻，或者是无法识别的原因的

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

影响。分析出现转折点的原因是股票分析师的工作。



图 2-8 调整后收盘价格轨迹（通用汽车），以 20% 作为转折点的识别标准

将克莱斯勒（在它被戴姆勒收购前）与通用汽车作为一个匹配交易，表 2-1 给出在几种不同事件序列下的一个对比简表。由于事件之间的回报是显著的，不同条件的时间序列，彼此之间存在差异是无关紧要的。后面这句话非常有用——识别事件的算法，校验规则是不是足够精确，对于事件与事件之间的相关性来说，影响并不是很大（也就是并不那么敏感）。因此，在相关性分析中，不需要过分关心选择哪一组事件才能选出风险控制良好的匹配交易。

在成交量序列上，使用事件分析法，有时能识别在价格序列（使用转折点分析方法）中不能识别的信息。成交量模式不能直接影响价差，但是成交量急剧扩大的现象，是一个很有用的警告。这种警告信息通常表示股票可能受到不寻常的交易活动的影响，价格的发展可能会不再符合由最近历史价格序列平均值估算出来的统计模型。在历史分析中，标识异常变化，对于估值（如模拟结果）来说是非常重要的。识别历史数据中成交量的峰值，跟之前如何识别历史价格的峰值，几乎一样。然而在实时的交易中，监测成交量

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

的增加情况，是一项重要的风险管理工具。我们要在成交量峰值被识别出来之前，在成交量增长期，识别交易的峰值，因为等事件发生之后才将它识别出来，那就太迟了。

表 2-1 克莱斯勒 - 通用汽车的事件回报简表

标准	事件数量	回报相关性
每日	332	0.53
30% 的移动	22	0.75
25% 的移动	26	0.73
20% 的移动	33	0.77

2.4.3 在 21 世纪的相关性搜索

现在，好几个供应商都提供匹配交易的软件，可以用来识别匹配交易的标的物和交易渠道，以及投资组合管理。这里指的相关性搜索，在 20 世纪 80 年代，就已经用程序实现了。同样的工作不需重复做。例如，瑞士信贷第一波士顿（Credit Suisse First Boston）就提供一个工具，让使用者将布林线类型的匹配交易模型，应用到任何特定的股票匹配交易中。这个程序用某个固定的时间周期，在一个均值加减标准差的模型中，搜索某个范围内的模拟交易，对不同的模型（不同的时间周期、不同的布林线宽度）进行比较，识别出产生最大利润的模型。任何人都能使用这样的软件，非常快地找到很多匹配交易，但仅仅依赖这样肤浅的数据分析是很危险的。高盛、Reynders Gray 以及其他公司，都提供类似的软件。

到目前为止，还没有出现非常容易使用的商业工具，去识别事件或转折点，以及计算事件之间相关性。

2.5 投资组合结构和风险控制

随着模型的发展，投资组合风险管理越来越受到重视。均值方差（mean-variance）方法在很长一段时间内受到人们的喜爱，被当成是一种将风险“控制在一定程度内”的方法。这样的一种想法，在 1998 年夏天被证

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

明是愚蠢的，在后面的章节会看到这个故事（见第8章）。

一些设计模型的人，会将风险暴露程度和回报预测整合在投资组合模型的程序中（见2.4节，以及第3章关于外部因素模型的相关描述）。其他人（特别是那些采用模糊的预测函数建立模型的人）是先建立一个投资组合模型，然后计算投资组合在市场因素中的暴露程度，再将这些暴露的风险单独进行对冲，以达到控制风险的目的。

目标是选择一组股票投资组合，让投入的资本获得最大的回报。如果有先见之明，最优的投资组合是将资本投入到能带来最大回报的股票中。当然，我们不能进行完美的预测。替代方法是，我们尽最大可能进行预测，然后做下一步工作。目标是将实际的回报最大化，在这个预测的世界中，我们必须将焦点放在期望的回报上。

预测不同于预言，不能保证结果会与预测相同。根据预测采取行动，必然有风险。预测在简单的匹配交易中，“价差将回归均值”，实际上价差可能会进一步扩大，只能采用止损方式清空头寸。考虑到风险这个新的因素，使目标变得复杂，而且要同时考虑两个目标：回报最大化和将风险控制在一个可以容忍的程度之内。

到目前为止，从预言到预测，我们从确定到不确定，从有保证的最优化到最优的猜测。然而，实际情况不是这么简单。第一个障碍是必须能精确地说明风险的定义，或者在实践中能应用，因为预测不一定能变成现实。事实上，如果预测100%的准确，这是一项极不寻常的事情。精确的预测只能产生一个结果。但是我们将唯一精确的结果，转化为无限多个可能的结果。很重要的事情是，预测必然会产生误差。

“根据最优的结果”变成了“根据最优的预测”，而且必须知道，最糟的情况会是什么，并且尽最大努力保证避免发生最不希望的结果。

与必须预测最优值一样，我们必须预测最糟糕的情况。我们猜测最糟糕情况的方法，与寻找最优的预测方法，有一点点不同：我们会重点考虑降临在自己身上的最糟糕事情的可能范围（从小到大），而不是寻找灾难将如何

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

发生（场景分析通常用来提醒注意那些“不大可能发生的”，极端情况，而没有考虑投资组合建模的日常性的“风险控制”。极端情况和常规情况下的风险管理之间的区别被故意模糊化了）。

因此，我们的目标变为：将期望回报的波动率限制在一定范围内，使期望回报最大化。在方差的约束下，会使可供接受的投资组合减少。回报变动的期望值，必须低于某个值，才会采纳这样的投资组合。

有一件很重要的事情，所有的量（包括预测的回报和方差）都是不确定的。预测的方差，能指导我们存在多少个合理的预期结果。这些预测的结果与最优预测是存在偏差的。但预测回报的方差自身也是个预测值。它不是一个已知的数。请记住前两段所说的原理：预测的方差仅仅是对平均行为特征的描述，在任何情况下，什么事情都有可能发生。

在这么多警示性的评论之后，我们还会使用预测方式，因为这些方式具有一些预测作用。一般说来（对特定情况不成立），采用预测方式预测未来的事件，比随机预测要好得多。同样的，可以运用预测的方差来量化预测结果变动的范围。

最后，我们对上述方法进行实际操作，并量化风险。我们将投资组合的风险定义为投资组合方差的期望值。我们对风险的厌恶程度可以描述为一个常数乘以方差。这样，我们的目标变为：在一个回报方差期望值给定的情况下，使回报期望值最大。

让我们用数学公式来描述这些结果。首先，定义一些变量：

n——投资股票数目

f_i——股票 i 的预测回报的期望值；$f = (f_1, \cdots, f_n)'$

$\sum$ ——回报方差的期望值，$V[f]$

i_p ——股票 i 的投资价值；$p = (p_1, \cdots, p_n)'$

k——可容忍风险程度因子

目标函数表示为：

$$\text{maxmine} \quad p'f - kp' \sum p$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

2.5.1 市场风险因素的暴露程度

一般情况下，统计套利基金经理不会采用多方头寸的投资组合：这样的投资组合直接暴露在市场风险中（根据定义，一个匹配交易方案不会偏向多头或空头，但是统计套利模型一般都会产生预测，如果不加以限制，常会得出一个倾向多头或空头的投资组合）。如果市场遇到崩盘的情况，投资组合的价值也会一起崩盘。这通常与投资组合的组成项目没有什么关系。需要建立一个市场中性的策略，目标是在市场发生波动时，能找出一种让股票不会随之波动的方法。这个问题与如何界定“市场”有关。根据一般的规则，标准普尔500指数通常被视同为整个市场（或市场的代表）。都用统计的方式，检验投资组合中的每一只股票，对每一只股票暴露在标准普尔指数的风险程度进行量化。这些量化的结果用于测定投资组合暴露在市场风险中的程度。要达到对市场中性，可以通过改变投资组合中股票所占的比重来实现。

定义如下：

l_i：股票 i 暴露在市场中的风险程度；$l = (l_1, \cdots, l_n)'$

投资组合 p 暴露在市场中的风险程度：

$$\text{暴露在市场中的风险程度} = p'l$$

为了达到市场中立的目标，将目标函数修订为：

$$p'f - kp'\sum p - \lambda p'l$$

其中 λ 是一个拉格朗日乘数（Lagrange multiplier，仅仅与优化策略有关）。

可以将对市场中性的追求，从整个市场扩展到某个行业中去。例如，我们想避免暴露在石油产业中的风险，保持市场中性，在做法上是相同的：定义每一只股票暴露在“石油行业”中的风险。应该注意到：这种方法比先为石油行业定义一个指数，然后计算每只股票在该指数中暴露的风险程度相比较，是一种更通用的方法。对于每一只股票，不管是否属于石油行业，都在石油行业中暴露一定的风险。如果给出的一系列风险暴露的状况，目标函数可以用相同的方法进行扩展，考虑这些市场因素。

定义如下：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

$l_{1,i}$ 为股票 i 暴露在石油行业的风险程度；$l = (l_{1,1}, \cdots, l_{1,n})'$

将目标函数扩展成为：

$$p'f - kp'\sum p - \lambda p'l - \lambda_1 p'l_1$$

其中 λ_1 是另外一个拉格朗日乘数。

很显然，其他市场风险因素可以放在目标函数中，保证投资组合在其中的风险暴露程度为零。如果有 q 个市场风险因素，目标函数可以表示为：

$$p'f - kp'\sum p - \lambda p'l - \lambda_1 p'l_1 - \cdots - \lambda_q p'l_q$$

求投资组合目标函数最大化的方法，可以直接利用拉格朗日乘数方法。

2.5.2 市场冲击力

我们预测 IBM 的股票在下周将有 10% 的年度收益率。假设这个预测比其他的任何预测都要准确。我们想购买 1 000 万美元的股票。一般来说，这么大的需求无法全部用当前的价格成交；最有可能的情况是，随着市场价格被拉高，购买需求逐渐得以满足。这就是市场冲击力，不管交易量的大小，大多数交易都会对市场产生冲击力。这是因为市场并非静态的，从价格的预测到最后成交，市场价格呈动态变化。没有实际交易数据，是不可能估计对市场的冲击力。即使有了交易数据，也只能是可以进行预测了：我们又要对一个不确定的事件进行预测（统计套利在这方面的进展情况见第 10 章）。

市场冲击力具有非常重要的意义。如果能够准确估计交易的成交价格，就能使交易系统从投资组合优化方式中，过滤掉无法获利的交易。

很快就出现一个问题：在建立预测函数时，是否能将市场冲击力考虑在内？认为可以的想法是很肤浅的。在没有特殊说明的情况下，股票可以用当前价格成交。好，既然如此，那么为什么完成一个交易时，时间上的延迟会导致成本上升？难道我们不能预期，有些成交价格将会有利于成交，有些价格则不利于成交，多个交易日的大量交易的相互中和，就会产生相互抵消的效果吗？这种想法也是肤浅的。在建立模型的过程中，我们对市场的影响并没有考虑在内。我们的买入订单会增加需求，拉升价格；卖出订单会产生相

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

反的效果。因此，我们自己所进行的交易，会在市场中形成对我们不利的力量。所以，我们的预测只有在我们没有采取行动的情况下才是正确的。

有人会问既然是建立一个可应用的函数，为什么不将这个交易也纳入到建模过程中综合考虑？简单地说（其实不简单），这么做实在是太困难了（换句话说，所需要的数据是无法获得的，目前能做的事情很少，至于能做到什么程度见第10章）。因此，比较务实的方法是，我们先将自己置于一个消极的观察者的位置，去建立一个有效的预测模型，然后再考虑如果我们积极参与其中可能会产生何种影响，继而对模型进行调整。

市场冲击力是我们进行交易决策的一个函数。我们用 c 来表示目前的投资组合，那么目标函数可被扩展为：

$$p'f - \text{市场冲击力}(p-c) - kp'\sum p - \lambda p'l - \lambda_1 p'l_1 - \cdots - \lambda_q p'l_q$$

对于大多数市场参与者来说，求解“市场冲击力”函数是一个尚未解决的课题，其困难是数据不充分（一般来说，受到交易者自己的交易单与订单的限制），而且在某些情况下，也缺乏解决问题的动机。同样的，关于这方面的发展也见第10章。

2.5.3 利用事件的相关性控制风险

在上一节中，我们探讨了有关事件相关性的一些观点，将它视为在股票池中识别共同风险因子的基础工具：几只股票重复出现几个相同方向的价格变化，而且在这几个变化之间，变化量也很类似。在同一组股票中，对于新闻或者所谓的“事件贝塔值（event betas）”，会具有相类似的反应。

如果利用一组具有类似事件贝塔值特性的股票，建立一个对事件匹配的投资组合。这个组合不管是多头还是空头，都对事件具有相同程度的反应，这个投资组合具有一定的风险控制能力。每一组投资组合都包括一些在经济发展中反复出现相同价格波动的股票。这个关键的特征是重复性的变化。正如在电影《007之金手指》中，奥瑞克·金手指（Auric Goldfinger）对邦德说的那句话：“一次是偶然，两次是凑巧，第三次就是危险在向你示意了。”当市场受到冲击时，投资组合中的股票不会出现完全不同的变化，受到市场

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

影响的概率会比较低。2001 年，恐怖分子袭击美国之后，虽然市场出现了灾难性的下跌，个别股票波动率也出现剧烈的变化，但是“事件贝塔中性”的投资组合只显现出较小的波动。

2.6 动态变化和校验

反转现象模型会用到局部的数据。例如，计算股票之间价差的波动率，要使用加权的方式对局部的数据进行赋权（见第 3 章）。在本章前面用到的校验方法中，我们以 60 天作为一个周期，然后从以下二者中选择其一：①直接使用价差的范围值；②使用价差分布的标准差。这些每天都更新的值，会不断地调整目前交易的进场和出场的时点。采用类似的做法，我们每天都更新流动性估值，修正交易量的大小，以及投资组合的头寸。因此，模型本身没有改变，会根据局部的市场条件，进行一连串的自我调整。

模型偶尔会进行重新校验（或者管理者重新检验模型）。回顾一下大陆航空和美国航空价差的例子，价差的范围戏剧性地从 2000 年的 20 美元变为 2002 年的 6 美元。

对于一个利用反转现象的模型来说，我们也可以利用调优运算技术，揭示反转现象的持续变化特征。股票价差呈现出反转现象的概率很大（参见曼德尔布罗特的分形分析）。模型的设计者通常会选择其中一个概率作为校验的目标，选择的依据包括参数值的稳定性、建模人员的偏好、研究成果以及个人运气等。应用调优运算，也会监控其他的几个概率（模型校验），以体现不同概率所对应的变化信息。当然，其中总是少不了会有噪声，这些噪声每月都存在，不管在实际交易中，还是在模拟中，噪声在一个概率中的表现，与其他的概率没有关系。了解噪声这种现象是很关键的，在交易模型（模型空间中）与另外一个相近的竞争模型相比较时，如果出现比较差的表现时不至于就将它解释为需要对模型进行修正。这种方法还有助于模型的进化。有好几年的时间，某些特定概率的模型表现出很好的绩效，而这种表现说明不同概率的模型对于能影响反转的趋势。如果我们能将它识别出来，就

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

可以对交易模型进行校验和修正。

我们使用一阶动态线性模型（见第3章）分析一个经典的匹配交易策略，会发现只要持有一只大额股本的股票两周的时间，就能获得惊人的收益。在2000年3月，业界发现一种始于1996年的低频率策略趋势。这种趋势的信号首次出现在1996年，但变化的幅度还在局部的变动范围之内，这一信号很久以后才识别出来。1997年的反转移动还是能获得收益。但到了1998年，由于信用评级和长期资本管理公司造成市场崩溃，使统计套利绩效变得更加困难，充满风险。虽然我们可以发现一些信号，但我们并不相信这些观察结果。连续4年出现同样的暗示时，我们才开始意识到，不应该再把这种情况简单地归咎于预期中的噪声干扰，这种变动趋势具有更重要的意义，我们应该认真地对待这些信号。在5年内，第一次重新校验“一直用于交易”的模型中的参数，而我们也期望这个校验能改善回报率，希望它能超过原来回报2～3个百分点。观察2000～2002年的模拟实验，我们发现，结果显然超过了预期，因为市场的发展造成了高频率交易策略绩效表现不佳，反而比不上低频率的交易策略。这方面的问题见第9章。

调优运算：单一参数举例

在图2-9的四个图形中，我们可以看到一个单一参数调优运算的例子。图2-9a显示了一个原型的反应曲线：针对模型系数可能的取值范围来说，这个操作策略对应的（模拟）回报，显示出如下的走势，先是稳定增加，然后到达顶点，最后迅速地下降。在反应关系不变的情况下，通过分析过去的数据，我们很容易找出让回报最大化的参数。

图2-9b说明了一些观察结果。每年的反应曲线都是不同的，有一些类似，但确实不同，这也是策略通常能够奏效的原因。当我们在执行一个策略，选定某一个参数值，需要了解反应曲线的形状和变动量，并且在可能的情况下，将二者与对研究现象的理解联系起来（这个例子中就是反转）。根据图2-9a求出回报最大化的参数，是一件风险很大的工作，因为反应曲线可能会发生一定的转变，以至于有可能使模型的表现一落千丈。在模型校验

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

阶段，风险管理也会出现这种情况，采用表现一般但风险较低的模型，而不是可能产生灾难性后果的模型。我们应该预料到，不可控的因素可能导致灾难性的后果，如果只是根据“可控”的因素，也不是一个合理的风险管理策略。

图 2-9c 显示了一个原型的调优运算反应曲线，反应曲线的平滑移动（反应形式本身会随着时间会发生变化）。在实际操作中，调优运算发生时，通常会与系统变化一起出现，如同在图 2-9b 中出现的情况一样。我们所看到的状况，就好像是将图 2-9b 与图 2-9c 中的曲线组合起来形成图 2-9d。

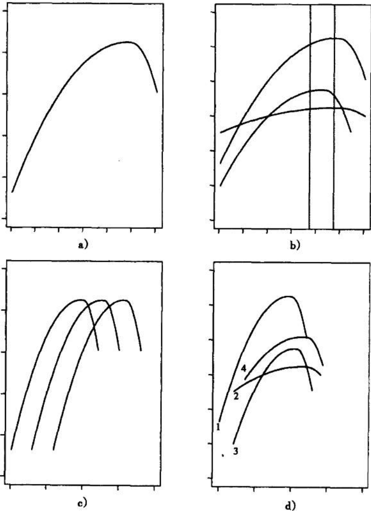


图 2-9 调优运算：检验持续的系统反应变化

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

由于反应曲线随时间发生变化，因此在参数中，前后一致的策略表现都不错，情况每年都不一样，而且随着时间的流逝，参数范围也会逐渐移动。或许原来的范围还是产生合理的表现，但是经过几年以后，就会越来越没有作用。调优运算就是通过反复监控模型的表现是否偏离了目前我们所认为的最优的校验值。哪些反应是暂时的，哪些是持续的。暂时的反应会提供信息，完善反应变动的观点。可以利用持续性的反应，改善模型的长期表现。

从上面的例子中，监测调优运算好像很简单，而且这个概念与旋转器的应用一样简单明了。不用怀疑，现实世界更加复杂。时间和变化的时点可能与这里以年为单位方式不同。当然，模型不是只靠一个参数，而是由一组参数决定。

以年作为监测的时间长度，具有一些主观色彩。变化有可能会在某个时点突然发生（如2001年9月11日），也有可能在一年内缓慢变化。从多角度对策略的执行状况进行监测，以便在不同概率下揭示诊断信息，是另外一项重要的工作。

统计套利模型有几个关键的参数。由于这些参数之间相互作用，所以监测方案更加复杂：一个参数变化所产生的冲击力，取决于对其他参数。因此，必须设计几个不同的模型绩效校验方案，并不断进行测算，以揭示策略绩效表现对个别参数的直接反应。

有些更复杂的模型牵涉到好几个分析步骤，可能包括数以百计甚至千计的参数。从概念上讲，监测问题在表面上看都是相同的：人们总是在寻找变化的证据，而不是出现在时间上的噪声。在实践上，采用较多参数的模型比采用少量参数的模型更加复杂，因为在包含诸多参数的模型中，很难解释每一个参数的意义。像“在参数 θ 中改变 X，意味着什么”这样的问题，很难回答。如果采用这样的模型，甚至问出这样的问题都很困难的。作为整体的组成部分，或者有时候是以一个层级结构的形式出现的一组参数，可能会具备综合的可解释性。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

在结束本章前，有一个重点需要重申：监测活动从它的机理，到解释说明，最后到采取的行动，都来源于对现象的理解，即它为什么存在？是什么产生了这个机会？以及在这个模型的背景下，模型是如何运作起来的？凡此种种，都是很重要的基础。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

第3章 结构模型

事实上，在现代社会，每一笔巨额财富的背后都隐藏着秘密信息。

——著名剧作家 奥斯卡·王尔德

3.1 导论

在第2章中，我们主要讨论了基于价差范围计算的交易规则，利用移动时间周期对历史数据进行计算得到价差范围。图3-1的折线图列示了大陆航空与美国航空的股票价差，计算方法以60天为时间周期，在平均值上加减一个标准差（我们可以将图3-1与图2-4进行比较，图2-4的上下限的计算方法是用最大值减去20%，最小值加上20%，其时间周期也是60天。同时可以回顾2.2节中的讨论内容）。这些交易规则中隐含了一个预测，即价差将会在不远的将来回归到均值。图3-2再次列示了大陆航空与美国航空的股票价差，不过这次图中包含了预测函数。

就未来的某一段时期而言，不管我们想要进行时点预测，还是计算预期值，一般的做法都是取目前的均值作为估计值。事实上，我们并不会真的认为每个时期的价差确实与均值相等，即使该时期就在不远的将来（很明显，交易规则所预期的是系统变量会围绕着均值上下波动）。以移动平均模型为基础对预测值的最优估计是，在不久的将来，价差可能会以均值为中心呈现

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

一系列数值分布。就像很多其他概念一样，至于多近才算是“不久的将来”，在这里没有详尽地说明。

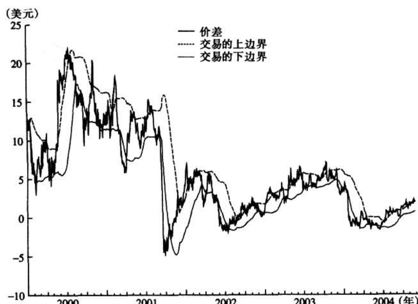


图 3-1　以标准差作为交易边界列示的大陆航空和美国航空的每日收盘价差

均值加减标准差这一交易规则，并不是通过统计模型得出的结论。相反，它是借助简单的目测加上略微的思考而推论出的一种假设。这种假设作为交易规则之后，在实际工作中取得了令人满意的结果。尽管如此，正如之前所说的，根据这些规则构成的模型，在预测函数上具有解释的作用。

在第 2 章中谈到，瑞士信贷第一波士顿银行的模型工具采用了比目测更进一步的方式，从诸多可选择的模型变量中进行系统搜索，包括对作为交易进出场边界条件的时间周期长度和标准差数目等。很明显，进行模型匹配或选择的流程是将效用最大化作为衡量标准，力求模拟交易利润最大化，而不是采用诸如最大似然性估计法（maximum likelihood，又称最大可能性估计法），或最小二乘法（least squares）之类的统计估计方法。效用最大化是一种复杂的方法，不幸的是，由于它范围受限，只是进行样本内交易利润的计算，因此在实际应用时效力大大减弱。事实上，以模型识别为目的建立的效

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

用函数在陈述上存在错误，可以预计，只要实际状况稍有不同，或许就会产生非常迥异的结果。真正令人感兴趣的是，将资本连续缩减的限度考虑在内，模拟交易预测规则的实际使用状况，探讨样本之外的部分能否实现利润最大化。但是，考虑这些情况之后，这一工具将朝着战略模拟模型的方向发展，而这样的模型是不可能免费提供的。

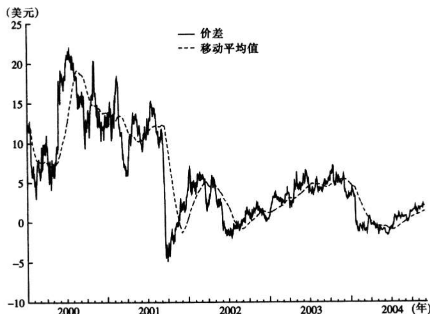


图 3-2 以移动平均值进行预测的大陆航空和美国航空的每日收盘价差

3.2 正式的预测函数

考虑正式预测函数的价值在于它给出了一组特定的数值，然后与现实情况进行比较，由此判断模型在应用效果上的功效。一项交易逐日发生损失暗示着可能存在潜在的问题；预测情况与实际结果之间的差异模式提供的有关信息，有助于说明问题的可能性质。这种信息认为同时关注多项指标对损失状况的反应情况，会比一个迟钝的止损规则（例如简单的损失百分比）更有效果。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

在本章中，我们将探讨一些最简单的、结构化的经典时间序列模型。我们也会介绍一两个非时间序列模型，阐释稍微复杂一些的对股价数据进行模型化的方法。

3.3 指数加权移动平均模型

移动平均模型（简称 MA），是我们从第 2 章开始就已经熟悉的一个概念。在此基础上，一个更灵活的局部序列平滑模型被称为指数加权移动平均模型（简称 EWMA）。MA 方法采用固定宽度的时间周期，并且全部数据都使用相同的权重，与之相对比，EWMA 方案则是对全部历史数据采用随指数递减的权重值。由此，越是近期的数据对当前的估计和预测的影响越大，而远期的重大事件则保持一定的权重。与 MA 的形式一样，估计 EWMA 的预测函数（设定 $k=1, 2, 3, \cdots, n$）也呈线性关系。不过，两者计算出来的值是不同的。对 EWMA 进行回归计算，其公式如下：

$$x_t = (1-\lambda)y_{t-1} + \lambda x_{t-1}$$

式中，y_t 是时间 t 对应的观察值，x_t 是 EWMA 的估计值，而 λ 则是主观贴现因子。主观贴现因子的大小介于 0 和 1 之间，即 $0 \leqslant \lambda \leqslant 1$，它决定了过去的观测值将会以多快的速度变得与目前的估计值不相关（换句话说，它决定了要取多长时间的历史数据计算目前的估计值）。

运用 EWMA 进行预测值计算所使用的回归形式，直观地告诉我们，EWMA 比 MA 的计算方法要简化很多。只需要在目前的预测值 x_t 的基础上，加入下一个观察值，就可以计算出一个新的预测值。虽然对电脑来说，无论是存储 1 个、20 个还是 50 个信息模块，其效果基本上都是一样的，但对人类而言，却存在着很大的差别。事实上，移动平均模型（MA）只需要取得两个信息模块，就能以回归的方式进行表达，从这一点可以看出，与 EWMA 相比，它在信息存储的需求支持方面就表现得更有效率。不过，EWMA 这种预测方法有其无可比拟的优点，下面就会加以举例说明。一旦熟悉了指数平滑法的预测方式，你可能就会把移动平均法束之高阁了。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

图 3-3 显示了使用 EWMA（0.04）与 MA（60）作为预测函数得出的大陆航空和美国航空的股票价差。EWMA 的贴现因子为 0.04，之所以选择这一数值，是因为其计算结果与 60 天移动平均的结果非常接近（贴现因子的选择采用的是目测的方式——其实可以选用诸如最小均方差之类的更严格的正式准则，但是，目前的情况并不要求采用这么正式的做法）。只有在原始序列（也就是价差序列）出现惊人的剧烈变化时，这两个预测函数才会出现明显的差异。表 3-1 列示了 EWMA 和 MA 在不同参数上的等效对照情况（前提是假设数据序列“表现得很正常”）。

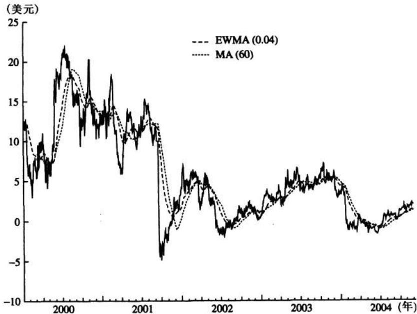


图3-3 分别用 EWMA 和 MA 预测函数得出的大陆航空和美国航空的股票价差

表 3-1 EWMA－MA 参数上的等效对照表

MA (k)	EWMA (λ)
10	0.20
30	0.09
60	0.04

在两种情况下，回归的方法会失效：一是价差遇到跳空变化，二是价差呈现出明确的趋势，而 EWMA 模型所具有的弹性则在解决这两种问题上发

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

挥了明显的效用。图3-4举例说明了一种情形，即价差突然变窄（价差值突然变小），之后会围绕着新的、更低的平均值上下波动。由均值和标准差所形成的区域（使用的是20天的周期）显示，在9月7日建立的多方交易，在价差突然减少之后，导致了高达11美元的损失，而最终这个交易，要一直等到10月11日，才会平仓（假设采用的是爆米花过程模型）。即使采用EWMA模型，而不是MA模型，两者在结果上也几乎是毫无差别的。不过，引进预测监视与干扰的概念后，EWMA在弹性上的优势就会显现出来。

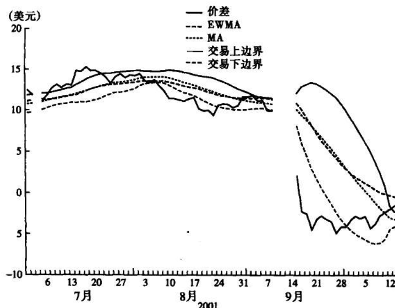


图3-4 价差水平的变化以及MA-EWMA的预测函数

当重大的预测误差发生时（例如价差出现不寻常的下跌），监视系统的触发点便会被启动，由此警告模型的设计者，出现了一个潜在的违反模型假设的情况，并会因此使预测失效。通过进一步调查，模型设计者有可能会发现价差出现这种行为的基本原因，进而去做出结束交易的决定（2001年9月17日那天发生的事情，其原因大家都很清楚，不过，在相同的情况下，是继续持有还是退出，对于当时的基金经理来说却是至关重要的，因为这直接关系到他们的业绩表现）。图3-5举例说明了一个预测函数，在该函数中，

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

没有使用价差的历史数据，而是用一个新的观察值取而代之。一般来说，预测的不确定性会非常大，图中很宽的上下限便说明了这种情况。

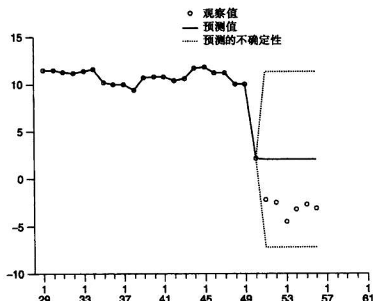


图 3-5 干预预测函数

如果没有发现新的信息，合理的做法是，密切关注接下来几天价差会如何演变（当然，还是要寻找基本的有影响的消息）。如果价差开始朝着之前的范围往回移动，那么就不需要采取任何行动。但如果价差继续围绕着新确立的水准移动，那么我们可以通过将这样的新认知引进到模型中，对预测模型进行改进。如果使用 EWMA 模型，进行调整就变得非常简单。只要针对一段期间增加贴现因子，赋予近期的价差更大的权重，预测值便会迅速地以新建立的水准为中心，如图 3-6 所示。通过这样做，我们能够更迅速地判断开始下注的价差值，以及了结旧交易，开始新交易的时机。在上面的例子中，在 9 月 21 日离场之后，虽然结果仍然是遭受了损失，但是资金会被释放出来，头寸风险也随之消失了。有利可图的新交易也很快会被识别出来。如果还是采用常规模型而没有进行调整，就会错失这些新的机会：9 月 27 ~ 10 月 2 日，以及 10 月 8 ~ 10 月 10 日，这两段期间，总共有 3.64 美元的获利机会。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

如果采用 MA 模型，同样也可以对预测结果进行监测，并对模型进行调整，但实际进行调整的时候，所牵涉的问题比 EWMA 难很多。对 EWMA 模型来说，只需要对贴现因子进行一次性的干预就可以了。试试看！

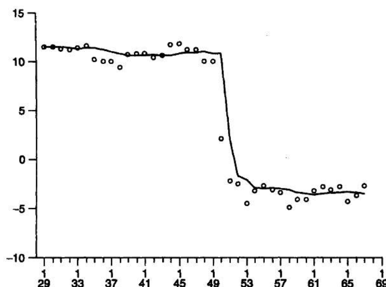


图 3-6　进行干预后的预测情形

新的贴现值是从哪里来的呢？在工程文献中，有许多复杂而吸引人的控制方案，但为了我们说明上的方便，这里只使用了一个简单的校验程序。首先，我们从一组价差的历史数据中，将价差发生变化的那个点挑选出来。接着，利用某个范围内几个不同的干预贴现因子，分别带入数据中进行试验，直到找出能涵盖所有情况的、适当的预测模式出现为止（再次提醒，“适当”是一个很主观的词汇，这需要你来解释）。

价差要出现多大的变化，才足以表示其在级别上可能出现了移动变化呢？回顾一下已有的数据：从平均值偏离三个标准差的情况，会多久出现一次呢？在进行校验之后，会产生多少次错误的监测预警呢？若采用四个标准差，情况又会怎样？有多少次级别变化被遗漏，而没有发出正确的预警呢？所采取的行动会对价差上的交易产生怎样的后果？考虑这些问题的时候，要注意的是，根据正态分布的标准差所推论出来的概率（例如，只有 0.2% 的

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

概率，会出现超过三个标准差的情况）并不是非常可靠，因为价差并没有呈现出典型的正态分布。我们不断地提醒这件事，但如果你想要更具体地看一下的话，可以将你偏爱的匹配交易的每日价差数据做成柱状图，然后将其与正态密度分布曲线进行最优匹配（套入样本的平均值和方差）。最后检查一下拟合的质量，尤其是在密度分布的尾部和中心部分。

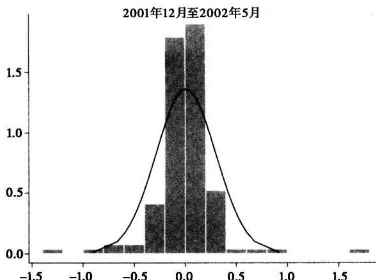


图 3-7 大陆航空和美国航空的价差正态密度分布的回报率柱状图

图 3-7 举例说明了一种普遍的情况。选取大陆航空和美国航空 6 个月（2001 年 12 月～2002 年 5 月）的价差作为样本，将其每日回报率以柱状图的形式列示出来，同时利用样本的均值与标准差，画出对应的正态密度分布曲线。这两者之间的差异情况，读者可以自己进行总结对比。

我们可以运用实证检验的方法，来拓展我们对市场的理解，这与采用正态分布假设相比，似乎是一个更加合理的方法。而且，如果你的目标是要建立一个具有代表性的模型，那么这种做法会是一个不错的起点。如果分布不是正态的，读者也不必担心。第 5 章针对时间序列数据的基础分布与回归之间的关系，澄清了一些常见的错误概念。

1994 年，在本人与其他几位作者的合著中，对以下三个问题进行了较

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

为详细的讨论：①对预测的监测；②干预与自适应方案，包括如何以似然性为基础的检验方式，而不是这里提及的较不成熟的标准差规则；③证据积累策略，等等。在那部著作中，我们将模型进行了详细分类，并将其统称为动态线性模型（dynamic linear models，DLM），MA 与 EWMA 模型只不过是其中的特例。而且，在一些统计套利者的交易方案中，呈现出显著的自回归特征的模型，也被包含在我们的分类中。与本书所讨论的内容相比，动态线性模型的结构提供了更丰富的分析框架。

在第 9 章第 2 节中，我们谈到了一个与此相关的实际发生的干扰状况：在 2003 年 10 月，受纽约司法部长调查骏利（Janus）共同基金从事择时交易活动（market timing，指基金违法进行短期的频繁交易）这一事件的影响，股票市场出现了巨额的基金赎回浪潮，并引致了规模高达 44 亿美元的抛售风波。而 2001 年 9 月，发生在美国的恐怖袭击所产生的预期市场反应，也提供了一个很好的例子，让我们意识到有必要认真检讨、仔细审视经过良好设计并精选的干扰因素的价值。

并非所有的评级变化，都像前面所提到的例子那样富于戏剧化。通常，价差会在经过几个交易日的移动之后，才到达一个新的评级，而不是由一次性的、巨大的跳跃来实现的。图 3-8 显示了英国石油公司（British Petroleum，BP）与荷兰皇家壳牌公司（Royal Dutch Shell）的价差，从图中，我们能看到多次这样的缓慢移动之后的突变。图中列示了两条 EWMA 预测函数曲线。第一条是标准的 EWMA 曲线，其贴现因子为 0.09（类似于 25 天的移动平均，不过在遇到明显的变化时，MA 对于数据的调整反应，要落后于 EWMA，前文已经述及）。在 2003 年第一季度，当价差级别发生 4 美元的移动变化时，该条 EWMA 曲线呈现出一个非常缓慢的适应过程。第二条 EWMA 曲线，则是在觉察到级别平移开始出现时，就调整到一个更高的贴现因子值。我们可以明显地看到，在 2003 年 2 月，在价差向下平移的过程中，第二条 EWMA 曲线展现出更快的适应步伐，在随后的 4 月与 6 月，价差在向上平移的过程中，以及 7 月末向下平移的过程中，也都可以看到这样的快

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

速适应调整状况。

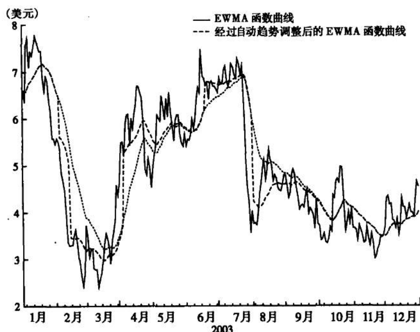


图 3-8 趋势侦测与级别调整

在这个例子中，使用的调整规则相当简单：以基本的 EWMA 为中心，只要价差连续几天超出正负一个标准差的边界值，就使用较高的贴现因子进行快速调整（局部标准差的计算，也是用 EWMA 的方式进行平滑化）。

在结束有关英国石油与荷兰皇家壳牌的价差讨论之前，让我们再回顾一下。在一整年当中，价差似乎围绕着 5 美元的均值，呈现出一种非常规则的类似于正弦曲线的变动状况。面对这样的情况，你将如何应对？

3.4 古典的时间序列模型

市面上有许多书籍，描述具有时间序列特征的预测模型。其中有一些列在本书所引用的文献中，这些书籍有助于读者理解模型种类、统计估计方法与预测流程，并对数据分析和模型构建能起到实际的指导作用。在本节中，我们将以启发性的方式描述几类模型，这些模型曾被成功运用到统计套利者

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

的实践中。讨论将会以前面提到一些议题为基础，包括有关价差、回归的描述，以及相应的原型化过程（如正弦曲线与爆米花过程）。

3.4.1 自回归与协整

自回归模型，可能是所有领域中，最常被应用的一种时间序列模型结构。这个模型的前提假设是，可以通过对时间序列中最近期间的数值进行加权平均，来预测序列的未来值。我们前面所提到的移动平均和指数加权移动平均，都属于这类模型。

所谓的 p 阶自回归模型，就是将时间 t 所对应的序列值的 p 个结果，进行线性组合，其表达式如下：

$$y_t = \beta_1 y_{t-1} + \cdots + \beta_{t-p} y_{t-p} + \varepsilon_t$$

其中，系数 β_t，或称模型参数，是通过对一组序列观测值进行估计后得出的。最后一项 ε_t，就是所谓的“误差项”，它可以方便地将那些影响模型，却又无法轻易地归入到模型的某一个组成结构中的方差因素，合并到这一单一项下。如果我们采用的是标准的估计方法，而且讨论的对象是参数估计值或预测值，那么，通常我们都会假定这个误差项将会遵循正态分布。

自回归模型，经常以 ARIMA 模型的形式出现，ARIMA 代表的是自回归整合移动平均模型（autoregressive integrated moving average）。在这一主题的参考书籍中，博克斯与詹金斯在 1976 年出版的著作堪称经典，其独特地位至今尚未被超越。在那本著作中，移动平均模型有点难懂，与我们之前讨论过的移动平均模型略有不同。实际上，它是应用一个数学等式，将一个很长的时间序列的移动平均值，以一种简短而晦涩难懂的方式表达为几个虚数项的平均值。理解这个公式让人颇费脑筋，所以，我们在此就不再对其进行深入探讨了。

那么，在 ARIMA 模型中，“整合”（integrated）指的又是什么意思呢？事实上，它只是我们在研究自回归模型的结构之前，对时间序列进行的一种差分运算。例如，我们用公式 $z_t = p_t - p_{t-1}$，明确表示每日的价格差异。如果差异值呈现出自回归结构，那么运用于原始价格序列的模型，就被称为

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

“整合”自回归模型。虽然从逻辑上看，我们之前介绍过的 EWMA 预测函数，最初是作为一种平滑观测变量的数据分析方法发展起来的，但实际上，我们可从这个函数中推导出最优预测整合模型。

由此，我们可以很顺利地引出协整的概念。通常，我们通过观察可以发现，几个时间序列似乎受到某种相互关联的关系驱动，呈现出共同运动的趋势；常见的情况包括：①其中一个序列驱动另一个序列；②多个序列同受共同的潜在过程所驱动。ARIMA 模型的多元表达形式，可以表示这类非常错综复杂的结构，包括，同时发生的情况，以及滞后反馈的关系。

其中有一个模型结构，应该是价差建模者比较熟悉的（但他们或许并不知道它有这样的一个技术名称），那就是所谓的“协整”。如果有两个（或多个）序列，各自都表现出非稳定的状况，但是它们之间的差异（在这里，我们称其为价差），却呈现出稳定的状态（可以将其意义理解为“用自回归的方式得到较好的近似结果”），那么，我们可以认为这两个（或多个）序列之间存在协整关系。换句话说，这些序列之间的差异（不一定是与第一个序列之间的差异，不过我们在这里不再对此深入探讨），可以用自回归模型进行恰当表达。

有一类与自回归模型相关的模型，可以为具有长期序列相关依赖性的时间序列，提供一个简化的结构形式。一般来说，我们可以通过使用一个非常高阶的自回归模型，直接求解这种长期的相关性，但是因为高阶模型采用了过多的参数，一些估计上的问题就会接踵而至。采用自回归部分整合移动平均模型（ARFIMA），可以克服参数过多的问题。本质上来看，这个模型就是先对时间序列做局部差分运算，然后再将 ARMA 模型应用到这个序列中。

3.4.2 动态线性模型

前面几节所讨论过的全部模型，其能否发挥效用，在很大程度上依赖于数据序列的稳定性。模型参数的估计，必须依靠一段相当长的历史数据，而且还要假定其应该保持不变。在金融实践中，对于任何一段时期来说，实际的数据“没有发生变化”，甚至连近似于“没有发生变化”的情况都很少

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

见。所以，存在广泛的对参数进行更新的方法，其中，将模型重新适配到移动的数据周期，是一种被普遍使用又很有效的策略。在第 2 章中，我们所用到的局部均值与方差的计算方法，就是一个典型的例子。

动态线性模型（DLM），是指一种灵活的模型结构，它在数据定义的特征上就直接体现了时间运动的特性（如局部平均值）。在动态线性模型中，模型参数会随着时间变化，这种变化是通过进化方程加以明确规定，从而被包含在模型中的。考虑下述一阶自回归模型：

$$y_t = \beta y_{t-1} + \varepsilon_t$$

在这个模型中，时间序列由两部分组成：第一部分，是由参数 β 所定义，表示对过去刚发生的数值，按照一个固定的比例进行的系统传播。第二部分，则是随机附加项 ε_t 。接下来，我们考虑将这一模型表述为更为灵活的一般表达式形式，其中，系统性的循序传递参数，在不同时期都可能会有些大小上的变化。现在的参数 β 将会与时间密切挂钩，不再是一个固定值，而其变动方式也受制于严格的形式规定，目的是允许这一参数的进化，但绝不能让其发生突变。这种动态模型，可由两个方程式加以表示，一种是定义序列观测值，另外一种则是定义系统性的参数演化：

$$\beta_t = \beta_{t-1} + \omega_t \text{ 系统方程式}$$

$$y_t = \beta_t y_{t-1} + \varepsilon_t \text{ 观察值方程式}$$

在系统方程式中，ω_t 是一个随机项，其方差大小，可以控制回归系数 β_t 的变化速度。如果 ω_t 等于零，动态模型就会简化为人们所熟悉的静态模型。如果 ω_t 具有“很大”的方差，历史数据序列就会立即大打折扣，而得到的 $\beta_t = y_t / y_{t-1}$ 这样的近似结果。稍后，以 3.3 节中的 EWMA 模型为例，你将会开始了解，动态线性模型中的干扰是如何得以实现的。

ARIMA、EWMA 以及回归模型等，都是动态线性模型的特例，因此我们可以将动态线性模型当成是一个丰富的、具有灵活分类且能够奏效的模型。我们可以轻易地定义监测与干扰策略，无论是针对模型的各个组成部分，还是把它们作为一个整体考虑。关于这方面的例子，可以参见 1994 年本人与

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

其他几位作者共同出版的作品。

3.4.3 波动率的模型

在数量金融领域，波动率的模型化，可谓渊源很深。虽然它在统计套利领域的运用，不如金融衍生品定价领域那么直接，在金融衍生品定价领域中，我们可以看到很多理论发展与公开应用，但波动率的模型化对统计套利来说同样是非常有用的。在本书中，我们将考虑的范围仅仅局限在之前已经进行过相当多讨论的简单价差模型中，其中包含：①原始数据的方差，该部分决定了回报的大小，以及潜在的交易机会数量的多少（这关系到一个策略中，备选原始数据能否经受住基本的生存考验）；②交易存续期间，逐日的损益变动情况（牵涉到一个策略的风险预测、止损规则）；③由随机共振（stochastic resonance）所衍生出来的收益（参见3.7节）。

在波动率模型化中，广义自回归条件异方差模型（GARCH）与随机波动率模型，具有非常重要的地位。在衍生品文献中，我们可以看到从基本的GARCH模型衍生出来的各种模型的变体，其范围（在此我们只取首字母缩写）从 AGARCH、EGARCH 到 GJR GARCH、IGARCH、SGARCH 和 TGARCH。GARCH模型的结构是在线性回归模型的基础上，加上一个误差方差项，用以代表非线性的影响因素。这个误差方差，在其他的模型中被假定为一个常数，或是一个与模型的某些变量或时间相关的已知函数，具体可参见本人与其他几位作者的作品中，有关方差定律的讨论。而在基本的回归函数中，这个误差方差，则被定义为误差项的线性函数。我们再看一下一阶自回归模型，这次把一阶GARCH模型作为组成部分纳入其中：

$$y_t = \beta y_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, h_t)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

其中，$\varepsilon_t \sim N(0, h_t)$，意思是假设误差项 ε_t 服从均值为0，方差为 h_t 的随机正态分布（“TGARCH”模型，采用的是学生氏 - t 分布，来取代正态分布，以平滑化模型对“大型”偏差的反应）。在这个模型中，预测值与相应的结果之间的差异，会被直接传递到随后的预测差异中。一项较大的预测误

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

差会施加一种“力量”，推动模型能够预测出即将到来的较大的波动。接下来，那也意味着，在更新参数估计值时，下一个观察值显得无足轻重。因此，如果将一个具有 GARCH 误差结构的模型，与呈现波动率聚类现象（爆发出比一般情况下更高的波动率）的数据进行匹配时，针对估计模型的不同结构部分，赋予变动比较大的观测值的权重，与变动相对较小的观测值比较起来，会在一定程度上被缩减。

一般的加权估计方法，是将方差的变异性，假设为一个已知的函数形式，（例如，经常出现在产品销售数据中的 1.5 次方这样一个水平，就是复合泊松分布函数的结果）。对比而言，GARCH 模型，则是将其作为模型结构的一个组成部分，来估计变量的变化模式。方差的变化模式，一般不会提前就被具体规定下来，但有一个规律：如果预测值与实际结果之间存在较大的差异，这往往意味着接下来将会出现比较大的波动率。

一些模型可能会包含更大的后缀结构，也就是带有更多的 ε_{t-k} 项，就像均值—方差结构中具有更高阶的自回归模型一样。要解释这样的模型存在一定的困难，而且并不令人吃惊的是，模型的成功应用要求较低的后缀结构，所以其应用也受到极大的限制。

有大量文献在探讨 GARCH 模型，最早可追溯到恩格尔在 1982 年发表的论文，文中谈及该模型在宏观经济学和金融学领域的应用。

3.4.4 模式挖掘技巧

直接通过模式挖掘程序，包括神经网络和小波分析等方法，来发掘持续的股票价格行为模式，这样的工作已经在着手进行中。小波分析是一种局部的傅立叶分析，它将一个时间序列分解成一群局部正交的基础函数，并赋予所考虑的原始序列以适当的权重。神经网络是一个加权变换函数的组合；虽然这些函数没有明确的时间结构，但是在输入（一个序列过去的观测值）到输出（预测值）的传递中，其实还是隐含着这样的结构。

神经网络是挖掘数据模式的一种优秀工具。只要同样的模式再次出现，网络的预测能力就会得到非常好的表现。不过它有一个严重的缺点，就是缺

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

乏可解释性。虽然在一个小型网络中（最多只有一个隐含层，每层都只有少数的节点，而且具有表现良好的传递函数），有可能理清这种传递关系，并由此得出理论上的解释，但是，这并不属于常规的状况。虽然如此，不过那又能怎样呢？如果神经网络能够成功地识别出股价数据中具有预测性的足迹，难么就算对于输入到输出的传递在思维理解上，不如其他应用因素残差项的自回归模型（见3.6节）那样严密，那又有什么关系呢？它可能对我们使用模型毫无影响，但是我们将这个可能引起争议的题目，留给您仔细思考。

神经网络最大的优势，就是它们非常具有弹性，而这也是他们从一开始就能够成为一种优秀的模式识别方法的原因。当结构发生变化时，神经网络可以非常快速地识别出正在进行中的变化，并随后对最新的稳定模式进行特征上的描述。但是，这样的弹性也总会伴随着一定的风险，那就是，被识别出来的模式可能只是存在很短暂的时间，就其可供利用的机会而言，它们的存在更是稍纵即逝。借用欧威尔（Owell）的一句话：“用来描述现状，没问题；但用其预测未来，行不通。”

3.4.5 分形分析

对于分形分析这一主题有兴趣的读者，我们推荐您去阅读该理论的创立者本华·曼德博（Benoit B. Mandelbrot）在2004年出版的著作，该书对有关内容的论述可谓十分精辟。

3.5 哪一类回报？

你想要预测的是哪一类回报？如果在你心目中，已经有了一个特定的范围，那么这一问题的答案似乎就显而易见了：先预测今天可能的回报，然后围绕着它进行交易。一般来说，如果没有限定一个包含理论前提、数据分析方法以及交易目标的范围，那么要找到这个答案就很困难了。让我们假设，交易的目标就是单纯地要实现战略回报最大化（当然，该目标也受制于一些

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

风险控制因素，在此我们对这一问题先不详加描述）。如果没有理论指导，我们可以只是去探究一些传统的时间维度，研究一下日回报、周回报或月回报的模式。更进一步的想法可能会建议你去研究一下回报在一定的期间内是如何进化的：检验 1，2，3，…，k 天的回报，可以表明一类特定序列的自然期间，无论这一序列是个别的原始股票价格，还是其中的函数因子（参见 3.6 节）；你也可能会想去检验一下接下来 m 天的最大回报。

有一些模式匹配模型，比本书所讨论的模型更复杂，存在更高的技术要求，该类模型引导我们去考虑，能否使用更具有普遍性的股票回报序列多元函数。

3.6 因子模型

迄今为止，关于模型化的讨论，主要集中在匹配交易的股票之间的价差上，因为统计套利所在的领域与匹配交易之间有着不解的渊源。现在，我们将讨论一些建模思想，这些思想被应用到具体的股价序列的综合分析之中。

“共同风险因子”这一概念，在 Barra 模型中被广为熟悉，它是所谓的股票收益因子模型的核心概念：其基本思想是，股票的收益可以被分解成两个部分，一部分是由市场中一个或多个潜在的因子（这一因子会影响所有的股票）所决定的，而另外一部分则是特定针对股票的，也就是所谓的非系统性收益：

股票收益 = 市场因子所产生的收益 + 非系统性因子所产生的收益

早期，在运用分解方式建立模型时，只是单纯地将市场因子等同于工业因子（例如标准普尔产业划分），和一个一般的市场因子。有些模型设计者，会使用指数来代表这些因子，为每日、每周或者每月的收益，建立起回归模型、自回归模型或者是其他模型；也有一种风潮，是根据①指数的预测模型，②结构性回归模型等，来进行预测，并据此建立投资组合。

之后，建模者也尝试使用一种更通用的模型，其被称为统计因子模型。在这类模型中，因子的估计值是根据股票的历史收益数据得到的，而股票的

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

收益可能与这些因子中的若干个相关，或者受这些因子所驱动。

3.6.1 因子分析

一组多元数据的因子分析，旨在对一个统计模型进行估计，在这个模型中，数据是由一组 m 个因子的回归来加以“解释”的。其中，每个因子本身都是可观察值的加权平均的线性组合。

因子分析与主成分分析（principal component analysis，PCA）有很多共同点，因为主成分分析更为人们所熟悉，所以可以将二者进行有效的比较。因子分析是一种以模型为基础的方法，而主成分分析并非如此。主成分分析关注的是一组数据，它在一定的观察空间中，找出那些揭示数据最大变动趋势的主成分。而因子分析是针对一组观察值的线性组合，也就是所说的因子，估计出它们所对应的权重值，目标是使观测值与模型拟合近似值之间的差异达到最小化。

虽然，主成分分析与因子分析之间的区别看上去似乎很模糊，但这并不重要。而事实也的确如此，我们不会对此再进行深入的讨论。

假设有一个股票组合，其中包含了 p 个组成元素（也就是股票）。我们可以有效地利用从某个固定日期开始的标普500指数的成分股，来作为举例的方向。指定因子的个数为 m，假定我们所选择的这个股票组合的每日历史收益是已知的，那么通过使用一个因子分析程序，我们就可以得到 m 个，由各自的因子输入量（系数）（*factor loadings*）所定义的因子。这些输入量会被赋予相应的权重，然后运用到组合的每一只股票中。这样一来，因子1对应的输入量为 $l_{1,1},\cdots,l_{1,500}$。而另外 $m-1$ 个因子的情况也是类似的，都有它们各自对应的输入量。

将这些输入量乘以各自对应的（并不是通过观察而得到的）因子，就可以得到组合相应的收益值。由此，给定输入量矩阵 $\boldsymbol{L}$，将刚刚描述过的输入量作为矩阵的列向量，而因子的估计值，或称因子得分（*factor scores*）就可以由此计算出来。

这样，在使用因子分析之后，我们除了可以得到原来的 p 个股票收益的

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

时间序列之外，还可以获得 m 个因子估计值的时间序列。把因子分析与工业指数结构进行比较，可能对我们理解前者是有帮助的；有些统计因子可能看起来很像工业指数，甚至可能直接被当做是工业指数。不过要牢记的一点是，二者之间存在重要的结构性区别，即统计因子只是股价历史数据的一个函数，其中并没有考虑公司基本面等有关信息。

如果将股票的收益回归到这些因子上，我们就会得到一组回归系数。对每一只股票来说，每个因子都会对应一个系数。这些系数揭示了股票受相应的因子影响的程度。就特定的因子构成而言，针对选定的评价准则（最常使用的是最大似然估计法），一组包含 m 个因子的观察值的线性组合，所能得出的最优回归结果是唯一的。但是，能得出相同回归结果的组合确可以有无限多种。采用不同的估计方法，我们则可以得到不同的最优回归结果。有些战略方法，诸如最大变异转轴法和主因子分析法（principal factors）——与主成分分析有关，可以被用来从无限多组的回归结果中，选出唯一的最优结果。

要注意的是，因子的输入量与股票受因子影响的程度之间，存在着二元性。这种二元性是因子本身定义与结构所产生的结果，而输入量矩阵的行向量代表的就是股票受相应因子影响的程度。也就是说，在 $p \times m$ 的输入量矩阵 $\boldsymbol{L}$ 中，元素 $l_{i,j}$ 对应的就是对第 i 只股票第 j 个因素的输入量，也是第 i 只股票受第 j 个因素影响的程度。

因子模型的含义该如何诠释？答案就是：一个包含 p 只股票的股票组合，可以被假定为，是由一组组更小的基本实体，即因子，经过大量混合之后产生的一个结果。粗略的观点可以认为，我们假设这个股票组合，实际上是受到一个被称为“市场”的因子所驱动的。而更精确一点的说法则是，我们可以假设，除了市场之外，还有一些“产业”因子也会对组合产生影响。那么，因子分析就可以被看做是，从整体股票组合所呈现出来的杂乱现象中理清头绪，找出因子结构的一种统计方法，进而指出我们所观察的这一股票组合，是如何基于“真实的”因子结构来构建的。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

3.6.2 剔除影响因子后的收益

另外一种基于因子分析的成功模型，则颠覆了常规的思维方式，其基本想法是：在构建一个预测模型之前，先把市场和产业划分活动这两个影响因子，从股票收益中剔除。这样做的逻辑依据在于：市场因素在一定程度上是不可预测的，但是个别股票所处的相对位置，在几天之内都会保持稳定，经过这样筛选出来的收益，应该会揭示出具有更强预测能力的结构。现在，让我们更进一步地探讨一下这个有趣的观点。

回归模型拟合（对影响股票价格的估计因素进行回归分析）之后，所得到的残差项，被称为剔除影响因子后的收益。为什么要将关注的焦点放在这些剔除影响因子后的收益上呢？我们的观点是，一只股票的收益，可以被看做是由两部分组成，即与一组潜在因素相关的收益（如市场、行业或是任何可能考虑到的其他解释性因素），再加上一些与特定股票相关的量。就股票 i 而言，其收益可以用代数式表示为：

$$r_i = r_{f_1} + \cdots + r_{f_m} + r_{s_i}$$

对于一个不受市场和行业因素影响的投资组合而言，与特定股票相关的组成因子，体现在标准拟合模型的残差项中。此外，与特定股票相关的影响因子，短期内可能会比其他影响因子项更容易预测，也更容易遵守这一模型结构。例如，不考虑对今天市场的总体看法如何，一组相关股票的相对头寸（值），很有可能与它们昨天的情况相类似。在这种情况下，基于“剔除市场因素”后进行的收益预测，我们构建一个投资组合，仍然有可能得到一个正收益。

一个简化的例子可能有助于我们理解这个概念所传达的本质。假设，一个市场只由两只股票组成，这两只股票在构成比例上大致相等（相似的股本，相似的价值，以及相似的价格）。t 日的收益可以被表示为：

$$r_{1,t} = m_t + \eta_t$$

$$r_{2,t} = m_t - \eta_t$$

在这个例子中，因子模型只包含市场因素这一组成成分，而且从历史数

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

据的分析中，我们可以看到，相应的市场值基本上与这两只成分股的平均值相等（更普遍的情况是，不同股票所具有的市场影响因子 m_t 是不同的，在该方程式中，m_t 的权重值与因子定义密切相关，这一点，我们在前文已经予以阐述，在此不再详谈）。在这种情况下，对于每一只股票来说，与特定股票相关的收益，将会表现出大小相等，方向相反的特征。现在，如果 η_t 这一变量，能被很好地预测，而不是随机猜测，那么不管市场收益模式如何，如果构建一个看多股票 1，并且看空股票 2 的投资组合（当 η 是负值时，情况相反），平均来看，我们将可以得到一个正收益。

在更现实的情况下，只要交易的数量足够多，而且剔除影响因素后的收益模型（如 EWMA 模型、自回归模型等），存在一定的预测精准度，那么你所采取的交易策略就会成功。如果根据预测收益的大小决定所下交易的大小，并对所选择的投资组合进行风险优化，那么我们应该可以改善模型的收益/风险状况。

3. 10 节将简要探讨因子分析的代数式表达，并对外部影响因子后的收益模型的结构进行讨论。

剔除影响因子后的收益模型的运行结构因子负荷值或风险暴露程度，一定要被定期更新，目的是使预测（以及在此基础上进行的交易）的绩效可以维持在合理的预期范围之内。因为统计因子直接反映或假设了历史股价的结构，所以发现这种结构呈动态变化并不令人感到惊讶。因此，根据过时的数据，来估计因子之间的关系，所得到的预测模型的结果，八成不会有好的表现。类似于我们之前讨论过的动态模型的组成要素一样，选择因子更新的频率，是跟研究者相关的一门艺术。每季或每半年调整一次，是经常被使用的循环周期。

计算剔除影响因子后的收益，一定要使用最近期间刚发生的一组负荷估计值，而不能使用同期产生的一组值，这样才能确保剔除影响因子后的收益序列，总是能将影响因子剔除于样本之外。虽然这样做对于模拟结果来说是有害的，但就执行战略而言，却有着非常关键的意义。这种普遍公认的做法

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

说起来很容易，但是，在一个复杂的模型或估计方法中，它却常常容易被忽略掉。

另外，我们可以考虑一个动态模型，它是 DLM 的一般形式，这个模型的参数根据一个结构方程式逐日加以修正，不过，由于这种做法在计算上非常复杂，所以，在 20 世纪 80 年代晚期的时候，在实践上还很难证明其合理性。今天，复杂的计算已经不再是一个难题，而且应用于股价预测的动态因子模型也已经出现在相关的统计文献中了。有了这些复杂的模型，允许操作上更大的弹性变得极为容易，而且也充满了诱惑性，人们会不知不觉地去采用只遵循数据的模型路径，而不再做其他考量。很多经理人最终都成为了这种诱惑的牺牲者，不知不觉地进行了一些优化的操作。

3.6.3 预测模型

对于每一只股票而言，在完成建立剔除影响因子后收益的时间序列所必须的全部工作之后，模型设计者仍然面临的一个问题是，要针对那些收益构建一个预测模型。那并非暗示，我们将收益打回了一垒的位置，因为如果是这样的话，就意味着剔除影响因子化的逻辑依据是无效的。尽管如此，正如前面所述的那样，建模者必须面对构建预测模型的任务。例如，我们可以考虑自回归模型。要注意的一点是，在这里协整模型大概没有什么价值，因为在剔除影响因子的程序中，已经假设那些共同因子都被剔除在外了。

在这里可以考虑很多精心设计的做法。举例来说，在时间维度上，如果所选择的周期多于一天，我们就可以在因子估计中，获得更好的稳定性。

为因子序列构建预测模型并对相应序列进行预测，这是一种可供我们使用的非常自然的做法。但是，这种做法并不会取代剔除影响因子后的收益模型所能得到的预测效果：在前面一节所列举的简单例子中，因子预测相当于对 m_t 进行预测，或者是它的一种累积形式。被剔除的影响因子量仍然含在其中。

在考虑收益预测时产生的一个问题，至今尚未被解答，即：k 日之后的累积股票收益，与 k 日之后的因子估计值之间，存在着什么样的关系？按照

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

之前的讨论，另一个需要考虑的相关问题是：假定如果市场因素比被剔除的因子量还要无规则和不稳定，那么所做的预测就变得毫无用处了（因为如此一来，交易本身就会产生更多不确定的结果）。这些考虑表明，在一个有效的剔除影响因子后的模型中，要想获得收益，预测影响因子只是一项次要任务。然而，因子预测模型（不管其架构是否是剔除影响因子后的收益模型），在监测市场结构变化，以及识别这种变化的性质和程度方面，都是非常有用的。

3.7 随机共振

以模型为基础去理解价差或股价的时间动态特性时，我们要注意的另外一个关键问题是，在这一过程中，相关的分析能为我们阐释一些可以利用的可能机会。考虑一个带有爆米花过程特征的价差：这个价差在局部的时间范围内偶尔会背离它的“正态”均值，随后又会沿着一个被定义得相当良好的轨迹，向那个正态均值回归。这个正态值的水平并不是固定不变的。如果不是受制于某种能够产生诱导力量的行为的影响，诸如大宗交易这样的行为，那么价差便会围绕着一个局部的均值，呈现时大时小的震荡运动。这种运动是非常随机的——至少，在目前的上下文背景下，我们完全可以认为它是随机的。既然知道价差一旦“返回”到它的均值，基本上就会在均值的附近呈现出随机的变动，这就意味着，我们得到的建议是，可以对回归离场规则进行一些修改，把原来的“当预测值为零时就离场”，修改为“相对于入场点而言，当预测值运动到零的另一边并超过了一点点时，就可以离场”。这里“一点点”的标准所依据的是，分析在它下一次偏离均值（向上或向下）之前，最近一段时间围绕着均值游走所形成的价差的变动范围。在波动率预测模型、GARCH 模型、随机波动率模型或者其他模型中，运用这样的做法，应该都是有帮助的。

我们刚刚举例阐述过的在局部均值附近的随机游走，即“无干扰噪声”现象，就是所谓的随机共振。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

当你阅读前面的描述时，可能会有一种似曾相识的感觉。在预测值为零的活动期间，对围绕着均值的变动所进行的模型化描述，与价差变动的一般性描述，总体来说是非常相似的。按照数学家本华·曼德博 的理论，这种自我相似性（self-similarity）在自然界中随处可见。本华·曼德博为了研究和分析这种模式，创造了一个数学上的分支学科，称为分形学（Fractal）。2004 年，曼德博提出观点认为，分形分析可以为理解金融工具的价格运动提供一个更好的模型，这一模型要比目前数理金融学文献中所提到的任何模型都更有用。不过，我们尚不清楚，是否已经有任何成功的交易策略是运用分形分析建立起来的。曼德博自己也认为，他的工具对于金融序列的预测来说，还没有充分发展到具有完全的可行性的程度。

3.8 实践中的事情

股价运动预测的不准确性简直令人难以置信。尤其在你刚学完一门统计回归分析的标准入门课程以后，更是要把这一点牢牢记在心上。传统的课程演示会宣称，如果 R^2 低于 70%，那么回归模型的作用就会大打折扣（一些统计学家甚至认为其毫无用处)。你如果没有学过这样的课程，也不知道 R^2 是什么，那么没有关系，就去阅读一下相关教材吧。传统课程所教授的内容并没有错误。只是相对于我们在这里所关心的情况来说，它显得不是很恰当。现在，观察一下你每周收益的回归情况，如果看到 R^2 的拟合结果等于或低于 10% 的话，你可不要灰心丧气啊！

如果想要成功利用不是很精确的预测模型，问题的关键就在于能够正确地预测股价变动的方向，并且其正确的概率要略高于 50%（假定无论预测变动方向是向上还是向下，其准确性都是一样的)。

［注意］现实状况更为错综复杂，因为它以一种“对基金经理更有利”的姿态出现。根据损益对称原理，我们可以陈述一个简单的观点，即一点小的偏向就可能驱动一个成功的策略：表面上看来，相对的胜算好像有利

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

可图，但实际上却很可能是噩梦一场。若干交易形成的集合最终得到的实际结果，是由利润总和减去损失总和决定的。一次大赢将与多次小输功过相抵。这一事实的重要意义在于，它引导基金经理去建立止损规则（如果一项交易没有按照预期的方向前进，那么就要提早离场），这样可以切断损失，同时又不会限制获利。表面上看起来，一个模型的结构只要相对而言有利于获利预测的机会更大一些，那么就可以在指定的风险容忍限度内，取得获利颇丰的交易结果。技术上来说，这样的规则通过运用一个程序来改变结果的特征，从而调整了模型的效用函数，而预测模型只是程序中的几大要素之一。

警告：千万要当心，别被那些随手可以拈来故事的金融商品承销商给愚弄了。如果告诉你，一个策略在正常情况下会产生一些小的损失，但偶尔可能会获得巨大的获益，那么，你一定会认为这种策略听上去十分诱人，尤其是在向你讲述了一个编织得非常巧妙的、似乎无法避免的灾难故事之后，这种策略就像救世主一样，使你以为找到了一个依照惯例胜算赔率总是偏向获利一方的灵丹妙药，它所展现出来的诱惑力，简直让你难以抵挡。在描述了剧烈波动的例子之后，总是会有一些具备较低风险（或较小损失），但较高获利潜力特征的替代方案，以满足市场需求的姿态被提供给投资者。这就好比使用稻草人进行伪装的技巧一样，很有说服力。或者称之为赫夫（Huff）所说的统计骗术。

精通这种话术的讲故事高手，会运用令人无可怀疑的概率微积分，为不可避免的毁灭，编织一个打动人心的故事。然后，便可以俘虏那些懵懂的人！只要用易于承受的小损失作为成本，就有机会获得一个意外的大胜利，这个故事就像蝙蝠侠拯救那些终日处于不知道“我到底该做些什么？”的日子中的人一样，这简直就是一个完全转败为胜的模式。就算不能（它可以吗？）转败为胜，但至少也不会瞬间一败涂地。相反的模式则一定会带来相反的情绪。听起来不错吧！

现在来看看那些所谓的小损失，其数量竟如此之多。在迎接蝙蝠侠带给

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

你的惊喜之前，先将这些小损失全部加在一起吧，你会发现经过一段漫长的时间之后，这些小损失会聚少成多，累积成一笔大损失，这种情况通常会被无耻地忽略掉（哦，我的意思是，会被不经意地隐藏在细节之中）。那么，我们真正得到什么了呢？在不可避免的突变发生之前，或许某段期间你会因为有所收获而感到欣喜，但很快，由于排挤效应，获利会拖延、溃殇引起消极心态、沮丧，绝望，而最后（如果你能坚持等待），或许你可以被无罪释放！到头来，它仍然还是一个充满了未知数的游戏，只是规则不同罢了。随机性会以各种姿态呈现在人们面前。

如果一个模型在 $(50+\varepsilon)\%$ 的时间中，能够预测出正确的变动方向，那么净收益就是交易的 $(50+\varepsilon)\%-(50-\varepsilon)\%=2\varepsilon\%$。只要下交易的次数足够多，这样的净收益就能够得以实现。“次数足够多”这一限制条款是至关重要的，因为只有在一个包含足够多交易的总体中，平均值才能被作为一个可靠的绩效指标。

只是保证一项战略最终的结果相当于你所下赌资的2%是不够的，就战略本身而言，若想确保其存在获利空间，所下的交易一定要充分考虑到交易成本。而且要记住，计算净利润时必须考虑在内的不是平均的交易成本。而是要考虑用净获利交易的很小的百分比除以更大的、所有交易的总成本。举例来说，如果我的模型获利的机会是51%，那么净利润就是交易的2%，即 51% - 49% = 2%。这样的话，平均来看，在100次交易中，获胜的机会是51次，失败的机会是49次。我每进行100次下注，就可以有2次获得净利。这在统计上是可以得到保证的结果。但是，我进行下注所需要的费用，是100次交易所需的费用，而不只是2次的费用。因此，我要保证获得2%的净利润，就必须支付全部交易的100%的成本才行。

统计预测模型的用途并非只是进行方向预测。它们也可以预测变化的大小。举例来说，考虑每周收益的一阶自回归模型：下一周收益的大小，可以精确地被预测出来（相当于上一周收益的一定比例）。如果一个估计统计模型具有一定的有效性，那么预测出来的那些大小就可以被用来改进交易的选

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

择：如果预测结果小于交易成本的话，它就可以被忽略不计。如果交易的期望收益小于其成本，就没有道理继续玩这场游戏，不是吗？

现在，我们来谈谈前面提到的所有预测的不精确性到底是什么情况？如果平均看来，这些预测还算可以，但个别表现却非常糟糕，那么我们又怎能够依赖个别的预测，去清除那些收益期望值低于交易成本的交易呢？难道我们能够保证不会将一些可获巨额利润的交易剔除在外了呢？更有甚者，反而将回报低于成本的交易留了下来？

事实的确有可能如此。我们在这里再一次谈到的，同样是一个与频率有关的论述。平均而言，被当成没有获利能力而遭到抛弃的交易，都是那些回报期望值低于交易成本的交易。而且平均来看，被保留下来的交易，都是回报期望值高于交易成本的交易。因此，统计上保证会获净利的交易，它的回报期望值都会高于交易成本[⊖]。

3.9 加倍交易：更深入的探讨

在对技术模型进行延伸讨论以后，甚至是在本章非常有限的讨论层次上，你都可能会被这些技术模型深深吸引，而忘记这些模型本身就是错误的事实。一些模型是有用的，但其有用性与我们使用模型的背景环境密切相关。当我们应用一个模型时，误差分析是一项很重要且必须要考虑的活动。知道一个模型在什么情况下会失效，又是如何失效的，可以让我们知道其弱点，进而发现如何去改善这些弱点并提高模型的有效性。



注 释

⊖ 我们应用配给（止损）规则到一个交易上，以期改进预测模型的结果。这样的程序增加了按照预测模型进行下注的平均收益，因此我们可以说，只要将条件限制得更严格一些，收益偏差有可能会把一些本来会蒙受损失的交易（平均收益低于交易成本的那些交易），转变成可以获利的交易。这是很精巧的方法，而且得到的结果也很美妙。关于这方面内容，也可以参考3.7节关于随机共振的讨论。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

在第2章中，我们介绍的“规则3”是一个加倍下注的方案，在这个方案中，如果价差变得足够大的话，就对这一价差加倍下注。这一想法是在一个已经公式化的模型，即规则1或规则2的背景下，通过观察模型中价差的运动模式而激发出来的结果。尽管这个类似非正式的规则是基于对数据的目测，而不是通过正式的统计模型建立起来的，但其实这也算是一种误差分析方法。

对于更复杂的模型，目测分析的方法是不可行的。那么我们必须明确的是，不管是在实践上计算在险价值时，还是运用综合仿真模拟时，都要将注意力放在模型的交易结果上。除了将预测回报与实际结果直接进行比较的标准方法之外，我们也可以检查交易的结果，看看从建立头寸的点开始，到解开头寸的点之间，交易结果运行过一条怎样的轨迹。在价差扩大就加倍下注的例子中，原始交易的累积收益所走过的典型轨迹是一条J型曲线：最初的损失会在后来得以弥补，然后，在交易加倍阶段，利润会逐渐累积。分析任何模型的交易轨迹，无论模型有多复杂，都有可能会显示出这样的演进模式。因此，这也为战略的改进，例如运用加倍下注方法，提供了原始的素材。

值得注意的是，揭示分析中暗藏机会的是交易的动态进化情况，而不是头寸建立与平仓时点的确认。动态进化、交易及其他一些问题是本文中反复出现的主题。虽然银行与投资者的报告最关心的是最终的结果，但对于问题与机会的识别来说，达成这一结果经历了怎样的动态过程才是最关键的。如果交易期间跨越了日历月份结束的界限，那么从解释月度收益变异性的视角来说，这一动态过程在帮助投资者进行理解方面也是很重要的。

图3-9显示了交易累积收益轨迹的三种原型：①从交易开始到平仓始终保持获利；②从交易开始到取消始终处于损失状态；③一开始损失，之后逐渐恢复并最终获利，形成J型曲线。对每一类中的大量交易进行分析，我们能够找到战略改进的可能性。想象一下，如果你发现历史价格与成交量具有某种明显的特征，并立即发出一种预示性的交易信号，这种信号将未来的交

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

易分成三类，这时候你该做什么[⊖]？不如自己想象一下。

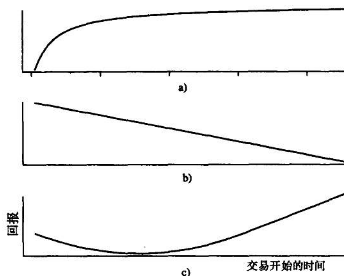


图 3-9 交易累积收益轨迹的原型

3.10 因子分析入门

接下来的材料，是基于莫里斯·肯德尔爵士（Sir Maurice Kendall）、艾伦·斯图尔特（Alan Stuart）和基斯·奥德（Keith Ord）的著作《高等统计理论》（现在被称为《肯德尔高等统计理论［KS］》）第 3 册第 43 章中有关因子分析的描述。这里用的符号是从 KS 修改而来，以便矩阵用大写字母来表示。因此，KS 中的 γ 在这里改用 Γ 来表示。这是为了保持本书在用法上前后一致。

假设有 p 只股票，其收益是由 $m<p$ 个无法观察到的因素值，以线性关


注 释

⊖ 这种类型的研究，在地震学中受到相当大的关注。在某些国家，预测地震一直是研究工作的重中之重，尤其是 2004 年 12 月，在印度洋发生了地震海啸，导致超过 200 000 人死亡之后，这类研究愈发受到重视。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

系决定的：

$$r_j = \sum_{k=1}^{m} l_{jk} f_k + \mu_j + \varepsilon_j, j = 1, \cdots, p$$

其中 ε_j 是误差项（观察值的误差，也就是模型结构中的残差项部分）。系数 l 被称为因子负荷值。变量的平均值 u_j 通常在分析之前会先被扣除掉。在本文的例子中，我们假设收益均值为零，因此 $u_j = 0$ 。以矩阵形式表示如下：

$$\underset{(p \times 1)}{r} = \underset{(p \times m)}{L} \underset{(m \times l)}{f} + \underset{(p \times l)}{\mu} + \underset{(p \times l)}{\varepsilon}$$

其中 L 是一个 $p \times m$ 的系数矩阵 $\{l_{ij}\}$（注意，这个表达式针对的是一组观测值，也就是，p 只股票在某个交易日的一组收益值）。现在假设：

（1）f 是独立的正态变量，其均值为0，方差为1。

（2）每一个 ε_j 都独立于所有其他的 ε，也独立于所有的 f，并且其方差（或特异性，*specificity*）为 σ_j^2 。

它遵循下面的方程式：

$$\mathrm{cov}(r_j, r_k) = \sum_{t=1}^{m} l_{jt} l_{kt}, j \neq k$$

$$\mathrm{var}(r_j) = \sum_{t=1}^{m} l_{jt}^2 + \sigma_j^2$$

这些关系可以用向量/矩阵的形式简单地表示为：

$$\Gamma = LL' + \sum$$

其中 $\sum$ 是一个 $p \times p$ 的对称矩阵 $(\sigma_1^2, \cdots, \sigma_p^2)'$ 。

从数据中，我们可以观察到 Γ 的经验值。我们的目标是决定因子的数目 m，并估计出 L 与 $\sum$ 中的常数。因子数 m 的决定具有高度的主观性：就像在主成分分析中，要先选定组成成分的数目一样。事实上，主成分分析法（PCA）经常被用来作为取得 m 初始估计值的一种方法，这样就可以利用似然率检验法（likelihood ratio testing），以及 m 因子模型的残差分析，来对它进行进一步的修正。在后面的讨论中，我们假设 m 是固定值。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

在某些情况下，在一个样本中特定某几天所隐含的因子分数值（factor scores）。也就是说，给定 t 日的回报 $r_{,t}=(r_{1,t},\cdots,r_{p,t})'$，对于 m 个因子来说，隐含值 $f_{,t}=(f_{1,t},\cdots,f_{m,t})'$ 是多少呢?

如果 $\boldsymbol{L}$ 与 $\sum$ 是已知的，利用广义最小二乘法，只要将下面的式子取最小值，就可以得到 $f_{,t}$：

$$(r_{,t}-\mu-Lf_{,t})'\sum^{-1}(r_{,t}-\mu-Lf_{,t})$$

要注意的是，平均股票收益向量（μ）被假设为零。（回想一下，μ_j 是股票 j 的平均收益：μ 并不是 t 日的平均股票收益。）将上面的式子进行最小化所得到的解是：

$$\hat{f}_{,t}=J^{-1}L'\sum^{-1}r_{,t}$$

其中 $J=L'\sum^{-1}L$。在实践上，如果 $\boldsymbol{L}$ 与 $\sum$ 是未知的；就可以用最大似然性估计法（maximum likelihood estimation，MLE）来取而代之。

在 10.3 节中，根据 S. J. Press 的《实用多元统计分析》（*Applied Multivariate Analysis*）一书，我们给出了一个估计因子分数值的替代方案（我们所使用的符号与此处相同，以保持本书前后的一致性）。基本上，这个方法假设不考虑误差协方差（error covariance），模型被重新描述成如下形式：

$$\hat{f}_{,t}=Ar_{,t}+\mu_{,t},\ t=1,\cdots,n$$

其中，在时间 t 时的因子分数值被表示成当时股票收益的一个线性组合。要求一个大样本近似值产生的估计结果用下式表示为：

$$\hat{f}_{,t}=\hat{L}'(nRR')^{-1}r_{,t}$$

3.10.1 剔除影响因子后的收益预测模型

在 3.6 节所描述的模型中，我们感兴趣的是剔除影响因子后的收益。对于 t 日来说，剔除影响因子后所得到的一组股票收益可以被定义为：一组收益的观察值与加权因子分数值（其权重当然就是因子负荷值）之间的差额：

$$\mathrm{d}fr_{,t}=r_{,t}-\hat{L}'\hat{f}_{,t}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

这个剔除影响因子后的收益向量是针对样本逐日进行计算的，它提供了用以构建预测模型的原始时间序列。举例来说，在一个自回归模型中，股票 j 在 t 日的回归项（entry in the regression）是：

$$\sum_{a=1}^{k} \mathrm{dfr}_{j,t-a} = \beta_1 \mathrm{d}fr_{j,t-k} + \cdots + \beta_q \mathrm{d}fr_{j,t-k-q+1} + \varepsilon_{j,t}$$

这个方程式说明，对于 t 日来说，k 日累积下来的剔除影响因子后的收益，会回归到 k 日之前，再往前追溯 q 个每日剔除影响因子后的收益（daily defactored returns）。要注意的是，对于不同的股票来说，回归系数都是相同的。

换个角度而言，在 t 日结束时，如果要预测接下来 k 日之后的累积剔除影响因子后的收益，则可以用下面的方程式进行表述：

$$\sum_{a=1}^{k} \mathrm{d}fr_{j,t-a} = \beta_1 \mathrm{d}fr_{j,t-k} + \cdots + \beta_q \mathrm{d}fr_{j,t-q+1}$$

你当然也可以采用其他预测模型：“你付钱，你决定。”

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net



第 4 章 反转定律

此时此刻，你必须明白，只有竭尽全力地奔跑，才能始终保持在同一个地方。

——英国著名作家　刘易斯·卡洛尔

4.1　导论

以本章为起点，我们将开始一系列旅程，用 4 章的内容，介绍在统计套利领域应用的价格波动的理论基础。本章描述的第一个结论，是一个简单的概率定理，它显示了一个基本的定律，证明在一个有效市场中存在价格反转现象。在第 5 章中，我们解释了一个常见但容易混淆的观点，指出即使价格分布出现“厚尾”现象，还是存在潜在的反转现象。总之，对于任意的数据分布，都有可能存在反转现象。在经过前述的解释说明之后，我们将在第 6 章中讨论股票之间价差波动率的定义和测量方式，以及在反转现象中扮演主要角色的波动现象。最后，作为整个理论系列的一部分，我们将在第 7 章中描述一个理论上的推论，即在匹配交易中预测将会发生多少次反转现象。

这 4 章结合在一起，举例说明并量化了在理想（或非理想）的市场条件下，存在着的统计套利机会。这些数据对于理解本书其他部分并非必不可少的，但是这些知识可以增强我们对市场发展的冲击力的鉴别能力。正是这种市场发展的冲击力，导致实践中统计套利在公共领域中表现失灵。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

4.2 模型和结论

我们描述了一个预测金融工具价格的模型，这个模型可以保证有75%的预测精度。我们所选择的假设是，对一组匹配交易的每日价差进行预测。如果对这个预测模型做稍微深入的思考，我们可以发现其非常广泛的应用。具体说来，我们会把焦点放在预测工作上，预测明天的价差值是否会大于或小于今天的价差值。

这个模型非常简单。如果今天的价差大于预测的平均价差，那么就可以预测明天的价差会小于今天的价差。另外，如果今天的价差小于预测的平均价差，那么就可以预测明天的价差会大于今天的价差。

4.2.1 75%规则

我们可以将刚刚提到的规则用公式描述为如下的概率模型。首先定义一个独立的、同分布的连续随机变量$\{P_t, t=1,2,\cdots\}$序列，这个序列落在一条非负的实线上，中位数为m。那么：

$$\Pr[(P_t > P_{t-1} \cap P_{t-1} < m) \cup (P_t < P_{t-1} \cap P_{t-1} > m)] = 0.75$$

用前面描述引发价差问题的数学语言来说，随机量P_t就是时间t这一天的价差（一个非负的数值），而每一天的随机量P_t都是独立的随机变量。这种概率的表达方式，是由两个组合事件所构成的，这两个事件与上面提到的非正式的预测模型所涉及的事件是一致的。但是对于每一个事件的细节来说，这些事件是有顺序的。这其中有一个很关键且必须要注意的事情就是，每一个事件都是同时发生的（且），而不是条件式的（如果……就……），尽管一开始就用“如果……就……”来描述非正式模型的特性。这个非正式的模型，可以说是我们采取行动的一个指南；而令我们感兴趣的是，那些（预测）行动的结果，其正确的概率有多大。因此，为了想要知道预测的表现，我们就必须要知道，在一个给定的时间，如果前一天的价差没有超过中位数，那么当天的价差超过前一天价差的概率有多大。同样，我们也想要知

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

道，在一个给定的时间，如果前一天的价差超过了中位数，那么当天的价差不超过前一天价差的概率有多大。

那些细微的区别看起来好像很神秘，但是如果想要正确的计算出一个投资策略的期望结果，那么了解这些细微的区别是非常关键的事情。假设 10 天中有 8 天，价差恰好等于中位数。那么也就表示，这个方案只能针对剩下 20% 的时间进行预测。我们必须要了解“且”与“或”的区别。如果出现错误的理解，那么就有可能会出现“预测的回报与实际回报之间比例是 5:1”这样的结果。

从操作上来说，一旦确认了今天的价差（交易收盘时），人们就可以对明天的价差结果进行交易。当观察到价差大于中位数时，就可以赌第二天的价差，会低于今天所看到的价差。在这样的一个方案中，赢得交易的比例可以用如下的条件概率公式表示：

$$\Pr[P_{t+1} < P_t \mid P_t > m] = \frac{3}{4}$$

同样，在另一个方向进行交易的人，也会有 3/4 的概率成为赢家。如果这样的话，那么我们有“1.5 的概率”获得赢的结果。而这就是真正的统计套利！在上面的例子中，没有考虑条件事件（conditioning event）发生的相对频率。现在，根据中位数的定义，只有一半的时间会发生 $P_t > m$ 这种情况。因此，在一半的时间里，我们会赌明天的价差低于今天的价差，并且这些交易有 3/4 的概率会赢。在另外一半的时间里，我们会赌明天的价差将高于今天的价差，而这些交易也有 3/4 的概率会赢。因此，所有的交易都有 3/4 的概率会赢$\Big($在前面的示例中，条件事件仅仅在 20% 的时间里发生，结果将是 $\frac{3}{4}\times\frac{1}{5}$，也就是 $\frac{3}{20}\Big)$。

在证明这个结论之前，要注意的是，“连续性”这一假设条件是非常重要的（在以后的模型表述中会重点说明）。在非连续变量的情况下，这个结论是不正确的，但是这无关紧要（见本节的最后一部分内容）。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

4.2.2 证明75%规则

证明这个结论要用到一个几何学上的论证方法，即将这个问题的结构用图形化的方式进行描述。这个方法还有个附加的优点，就是人们可以看到，在定理中某些结构性的假设是可以放宽的。在基本结论被证明之后，会进一步讨论这些假设放宽的条件。

在这里，考虑序列中两个连续变量的联合分布，也就是 P_{t-1} 和 P_t。假设两个变量是独立的，不管这两个变量的基础分布是哪一种，其联合分布的图形是对称的（对称于 $P_t = P_{t-1}$ 这条直线）。特别要说明的是，不需要假设，变量的分布必须是一个对称的密度函数 。

看一下图 4-1。在两个维度上，联合分布域（R^2 的正象限，包括零边界）都被中位数所在的点分割成两块。根据中位数的定义，这形成的四个象限分别各自代表了整个联合分布的 25% 的区域。

左下和右上的象限，以联合分布中位数为对称轴，被分成两个部分。现在，密度函数的图形呈现对称关系（结果来源于独立的和同一条件分布的随机变量），意味着每一个象限的一半，都有着相同的概率。因此，每一个半象限，都代表了联合概率的 12.5% 。

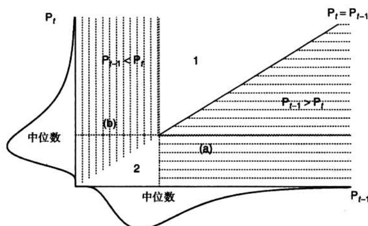


图 4-1 P_t 和 P_{t-1} 的联合分布域

剩下来要证明的事情是，将图形中对应的区域，与上面提到的概率描述

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

结合起来。区域（a）和（b）阴影部分的合集非常清晰和精确。这个合集也就是是整个联合分布区域，扣除了无阴影区域（1）和（2）之后剩下来的部分。每一个无阴影区域都是联合概率的 12.5%，这在前段中有所显现。因此，区域（a）和（b）的合集就精确地等同于联合概率的 3/4。

我们没有将左上或是右下的象限进行区分，因为这样做没什么价值。幸运的是，在一般的例子中没有特定的要求，因此没有必要对这两个区域进行区分。

4.2.3 75%规则的解析证明

采用几何学的论证方式，其目的是为了便于理解在下一节所探讨的一般化的结论。在继续往前证明之前，我们采用分析的方式，先提出结论。用 $X = P_t$ 和 $Y = P_{t-1}$ 来简化表达式。这两个事件就可以表示如下：

$$\{X < Y \cap Y > m\} \text{ 和 } \{X > Y \cap Y < m\}$$

表达式被拆分成两个部分（从 $Y > m$ 和 $Y < m$ 不能同时发生的事实中，就可以很轻易地看出来：在图形中，区域 a 和 b 不会重迭），因此每一个部分的概率，就可以简单地等于单个概率之和。考虑第一个部分：

$$\Pr[X < Y \cap Y > m] = \int_m^\infty \int_{-\infty}^y f_{XY}(x,y)\,\mathrm{d}x\mathrm{d}y$$

其中 $f_{XY}(x,y)$ 表示 X 和 Y 的联合密度函数。根据独立性假设，联合密度也就是单个密度函数的乘积，在这个例子中，X 和 Y 是同分布的变量（也是假设规定的）。通常用 $f(\cdot)$ 来表示密度函数，并且其相应分布则用 $F(\cdot)$ 表示，推导过程如下：

$$\begin{aligned}\int_m^\infty \int_{-\infty}^y f_{XY}(x,y)\,\mathrm{d}x\mathrm{d}y &= \int_m^\infty \int_{-\infty}^y f(x)f(y)\,\mathrm{d}x\mathrm{d}y \\ &= \int_m^\infty F(y)f(y)\,\mathrm{d}y = \int_m^\infty F(y)\,\mathrm{d}F(y)\end{aligned}$$

从最后一个步骤可以很容易得出，一个随机变量的密度函数是其分布函数的衍化形式。剩下的步骤不是很重要：

$$\int_m^\infty F(y)\,\mathrm{d}F(y) = \frac{1}{2}F(y)^2\Big|_m^\infty = \frac{1}{2}\left[(\lim_{t\to\infty}F(t))^2 - F(m)^2\right]$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

$$= \frac{1}{2}\left[1 - \left(\frac{1}{2}\right)^2\right] = \frac{3}{8}$$

对于解析函数的第二部分，在对一个初始的代数式进行简化之后，可以由一个类似的论证过程得出结论。首先要注意的是，事件 $Y<m$ 可以被表示为两个事件的合集：

$$Y < m = \{(X > Y) \cap (Y < m)\} \cup \{(X < Y) \cap (Y < m)\}$$

按照定义（注意 m 是分布的中位数），事件 $Y<m$ 的概率等于0.5。由于两个独立事件合集的概率等于每个事件概率之和，因此，可以利用这个定理将公式表示为：

$$\Pr[X > Y \cap Y < m] = \frac{1}{2} - Pr[X < Y \cap Y < m]$$

现在对第一部分公式的推导如下：

$$\Pr[X > Y \cap Y > m] = \frac{1}{2} - \int_{-\infty}^{m}\int_{-\infty}^{y} f_{XY}(x,y)\,\mathrm{d}x\mathrm{d}y$$

$$= \frac{1}{2} - \int_{-\infty}^{m} F(y)\,\mathrm{d}F(y)$$

$$= \frac{1}{2} - \frac{1}{2}\left[F(m)^2 - (\lim_{t\to\infty}F(t))^2\right]$$

$$= \frac{1}{2} - \frac{1}{2}\left[\left(\frac{1}{2}\right)^2\right] = \frac{3}{8}$$

最后将两个部分的概率相加，就可以得到最后的结果。

4.2.4 离散分布情形

假设一个随机变量只能取两个不同的数值，即 a 和 b，概率分别为 p 和 $1-p$。只要是 $p \neq \frac{1}{2}$，那么这个随机变量大于中位数的概率不会正好等于0.5。尽管这种情形不重要，并有些奇怪，但我们还是对这两个事件的概率定理公式进行研究：

$$\{X < Y \cap Y > m\} \text{ 和 } \{X > Y \cap Y < m\}$$

在这个离散分布的例子中，这些事件的概率公式可以表示为：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

$$\{X = a \cap Y = b\} \text{ 和 } \{X = b \cap Y = a\}$$

这两个公式的概率都是 $p(1-p)$，因此在这个定理中概率合计为 $2p(1-p)$。在 $p=\frac{1}{2}$ 的情况下（求导，令导函数等于零，然后求解），这个概率函数的最大值为 $\frac{1}{2}$。

4.2.5 普遍化

给出金融时间序列的数学表达式是很困难的（尽管曼德博在 2004 年时，对此提出了反对意见）。在统计模型中，这些数据特征包括非正态分布（在非正态分布中，收益常常被凸峰分布所影响）、非常数变量（市场的动态变化导致较高或较低的波动率，建立模型者也尝试过很多的方法，从 GARCH 到其衍生方法，到曼德博的分形分析等，具体见第 3 章），以及序列数据之间的相互依赖。针对上面所涉及的特征，可以通过放宽定理的假设条件来适应这些行为。

这个结果可以直接扩展到任意的连续随机变量：对于非负实数的限制条件不是必要的。在几何学的论证过程中，Y 轴（标识密度维度的轴）不需要精确的刻度（将原点选为零点只是为了便于说明）。在解析式的论证过程中，回想一下，我们并没有限制密度函数所覆盖的区域。

请注意，如果基础分布具有对称的密度函数（这暗示密度函数可能是一条完整的实线，或者是有限边界的区间实线），且密度函数存在，那么中心位置就是密度函数的期望值。柯西（Cauchy）分布有时适合运用于每日价格波动的数学模型。尽管柯西分布仅有一个中位数，结论依然成立。

在第 4.3 节中，75% 规则对非常数变量进行了进一步论证。

独立性假设条件也可以放宽：根据几何学上的论证过程，只要联合分布的图形是对称的，就可以放宽假设条件。因此，在定理中的独立性假设，可以用零相关性假设取代。在第 4.4 节中，用解析几何的方法和范例进行论证。

最后，对非常数变量的情形进行推广运用，每天的价差分布都可以有所

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

不同，使某些特定频率的假设条件也可以放宽。在第4.5节中，我们将具体说明这方面的细节。

4.3 非齐次方差

从一个给定的分布家族中，选出其中的一个分布，然后假定每天的价差，都是从这个分布中独立产生出来的。这个分布家族是固定的，但是某个给定日子，根据分布家族中的一个成员分布所产生出来的价格是不确定的。家族成员之间的区别只与变量有关。变量序列的特性也就决定了价格序列的特征。

如果每一天变量显现出独立的"随机"性，该怎么办呢？价差就会看起来不是根据分布家族的单个成员，而是根据所有家族成员的一个平均值产生，其中这个平均值可能来自变量的相对频率。换句话说，t 时刻的价差将不再是根据一个给定 σ 的 F 而产生，而是由分布的积分产生：

$$F_P(p) = \int F_\sigma(p)\,d\sigma$$

举例来说，假设变量的条件是一个正态分布家族（这里是指平均值为常数），而其变量是随机产生的，就仿佛是由一个逆伽马分布（inverse gamma distribution）产生的；那么，价差看起来好像是由学生 t 分布所产生的一样。这个结论的关键是随机元素，它保证（可能的）每天的数据看起来好像价差模型服从学生 t 分布（关于这点在4.5节中还会进一步讨论，一个类似的论证过程证明了，每天都采用任意产生的不同分布所得到的结果。边际分布与时间序列之间的关系在第5章中进行讨论）。

因此我们可以说，在不均匀的变量序列情况下，"75%规则"依然是正确的。

请注意，并不一定需要如前面例子一样，价差分布（变量是有条件的）和变量分布（无条件的）具有简洁的数学形式。任何一般的（"表现很好的"，从连续性的条件下）分布，都能得出75%的结论。密度函数或分布函

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

数，并不需要用一个封闭形式的数学表达式进行描述。

4.3.1 波动率异常变化

自动回归条件异方差模型（ARCH，恩格尔，1982）的产生，是为了在总体经济数据中，获得观察到的聚类方差（clustering of variances）的现象。在过去几年中，ARCH 和 GARCH 模型在数量经济学和金融学文献中应用非常广泛，其中后者在波动率异常变动的现象中经常被用到。大多数这种波动率异常变化，会使得波动率从一般的水平开始变大，比较有代表性的是，这种情况与公司的坏消息有关。从历史上来看，波动率变小的情况比较少。从 2003 年上半年开始，美国交易所的股票波动率出现变小的现象。在 2003 年和 2004 年，价差波动率达到了空前的低点；这种发展过程对统计套利的含义将会在第 9 章中进行检验。

当波动率异常变动时，标准差不是每天独立地生成，而是会呈现连续的相关性。根据下面的论证，75% 规则仍然成立：在波动率异常变动的过程中，如果与常数变量的状况（近似地）一样，定理也是适用的。因此，只有波动率出现波动的那几天，结论才会有所不同。这样的日子非常少，因此结论会非常接近于近似值。事实上，还可以这样证明结论：波动是不相关的（见第 4.5 节中关于一般非常数变量情况下的论证）。第 5 章将论述相关模式的分析。

4.3.2 数字上的示例

图 4-2a 展示了 1 000 个样本的柱状图，这些值是根据正态逆伽马（normal-inverse Gamma）分布产生出来的：

$$\sigma_t^2 \sim \mathrm{IG}[a,b]$$

$$P_t \sim \mathrm{N}[0,\sigma_t^2]$$

首先，根据逆伽马分布（有两个参数 a 和 b，a 和 b 的实际值并不是很重要：只要是任何的非负值都可以），生成一个独立的 σ_t^2；然后根据 σ_t^2 的值，以均值为 0、方差为 σ_t^2 的正态分布生成一个 P_t。在柱状图上，添加学

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

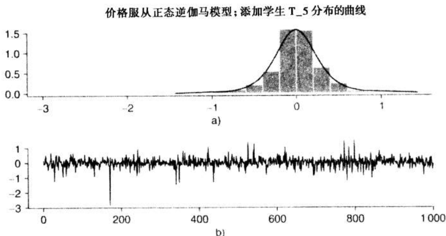


图 4-2 从正态逆伽马模型中抽取的随机样本

生 t 分布的密度函数，代表了价差的理论边际分布。图 4-2b 展示了一个时间序列的样本。

每天的价格变化向中位数方向移动的比例是 75.37%，与前面的规则一致。

4.4 一阶序列相关性

上面的结论可以扩展到变量具有相关性的情况。在最简单的情况下，可以考虑对称密度函数所对应的分布，因为在这种情况下，联合分布的图形不是圆形（变量是不相关的情况下）的，就是椭圆形的（变量是相关的情况下）。在后一种情况下，可以看到沿着图形的主要和次要的轴，将 R^2 分成几个象限，然后用联合中位数点，将象限以放射状的方式分割，从而又被划分成两个相同概率的区域（回忆一下，在对称的密度情况下，所有的象限都是按照等概率的方式划分，不仅仅对应到旋转坐标的左下和右上）。剩下的任务是用随机数量的方式，对这些半象限进行正确的描述（就像对原先的结论所采用的描述方法）。可以在范例中很容易地看到这个结论。假设 P_t 和 P_{t-1} 之间的协方差为 c。定义一个新变量，作为相关变量的一个线性组合，Z_t =

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

$a(P_t - rP_{t-1})$。系数 r 用公式表达为：

$$r = \frac{\text{cov}[P_t, P_{t-1}]}{\text{var}[P_{t-1}]}$$

（r 也就是 P_t 和 P_{t-1} 之间的相关性）令 P_{t-1} 和 Z_t 不相关；而因子 a 的选择方式，是为了让 Z_t 的方差等于 P_t 的方差：

$$a = (1 - r^2)^{-\frac{1}{2}}$$

现在将定理应用到 P_{t-1} 和 Z_t 上，假设 Z_t 和 P_{t-1} 具有相同的分布，因此我们可以得到：

$$\Pr[(Z_t < P_{t-1} \cap P_{t-1} > m) \cup (Z_t > P_{t-1} \cap P_{t-1} < m)] = 0.75$$

将 Z_t 进行替换，将公式转化为一个包括原先变量的形式：

$$\Pr[(aP_t - arP_{t-1} > P_{t-1} \cap P_{t-1} > m) \cup (aP_t - arP_{t-1} > P_{t-1} \cap P_{t-1} < m)] = 0.75$$

重新排列各公式项，得到需要的表达式：

$$\Pr[(P_t < (a^{-1} + r)P_{t-1} \cap P_{t-1} > m) \cup (P_t > (a^{-1} + r)P_{t-1} \cap P_{t-1} < m)] = 0.75$$

很显然，在不相关的情况下，也就是 $r = 0$ 和 $a = 1$，我们可以将其作为该情况下的一个特例。$Pt = (a^{-1} + r)P_{t-1}$ 的边界，按比例 R^2 的象限进行分割，其比例大小由相关性的大小决定。在不相关的情况下，象限一分为二，与我们之前看到的一样。图 4-3 显示了 $a^{-1} + r = \sqrt{1 - r^2} + r$ 与 r 之间的关系，当 $r = \sqrt{\frac{1}{2}}$，最大值为 $\sqrt{2}$（用一般的微分方法，令一阶导数等于零，然后求解）。

在这里的论证中，所采用的额外限制条件是非常重要的，要特别注意。假如具有边际分布的变量，在线性组合中保持原来的分布形式，这个定理能够适用变量之间具有相关性这种情况。对于正态变量、双变量的学生 t 变量，以及许多其他的情况，定量都是正确的。但是在一般情况下，定理并不是正确的。

当 P_t 和 P_{t-1} 完全相关，也就是在（$r \to 1, a \to \infty$）的限制情况下，原来的结论就被打破了。失效的原因是由于奇点（singularity）的存在，因为原先自由度为二（两个有不同的日子或观测值），在奇点上其自由度为一（两个不同的日子限定成具有相同的价格，因此定理的反转是一种不可能发生的情况）。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

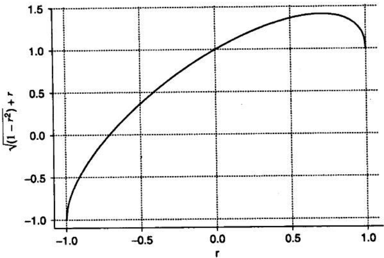


图 4-3 r 对比 $\sqrt{(1-r^2)}+r$

4.4.1 解析检验

一天之内的价格向中位数方向移动的频率，可以用概率的方式描述：

$$\Pr[(P_t < P_{t-1} \cap P_{t-1} > m) \cup (P_t > P_{t-1} \cap P_{t-1} < m)]$$

考虑上述公式的第一个部分：

$$\begin{aligned}\Pr[P_t < P_{t-1} \cap P_{t-1} > m] &= \int_m^\infty \Pr[P_t < P_{t-1} \cap P_{t-1} = p]\mathrm{d}p \\ &= \int_m^\infty \Pr[P_t < P_{t-1} \mid P_{t-1} = p]\Pr[P_{t-1} = p]\mathrm{d}p \\ &= \int_m^\infty \Pr[P_t < p \mid P_{t-1} = p]\Pr[P_{t-1} = p]\mathrm{d}p\end{aligned}$$

这里符号被广泛使用，仅仅是为了强调逻辑关系。对于连续变量，$\Pr[P_{t-1} = p]$ 是不正确的，因为这个变量取任何特定的值时，其概率都为零。正确的表达式，应该是在 p 点上计算出来的密度函数：

$$\Pr[P_t < P_{t-1} \cap P_{t-1} > m] = \int_m^\infty \Pr[P_t < p \mid P_{t-1} = p]f(p)\mathrm{d}p$$

请注意，当 P_t 和 P_{t-1} 是独立的，条件累积概率（第一项）转化为无条件的概率值 $\Pr[P_t < p]$，这种情况在 4.2 节中进行了描述。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

在接下来的结果推导中，对符号进行简化，用 X 来表示 P_t，用 Y 表示 P_{t-1}。那么，前面提到的概率公式可以转化为：

$$\Pr[X < Y \cap Y > m] = \int_m^{\infty} \Pr[X < p \mid Y = p] f(p) \mathrm{d}p$$

将条件累积概率 $\Pr[X < p \mid Y = p]$ 展开，将其放到条件密度的积分公式中，得到：$f_{X|Y}(p) = \Pr[X = x \mid Y = p]$

$$\begin{aligned}\Pr[X < Y \cap Y > m] &= \int_m^{\infty} \int_{-\infty}^{p} f_{X|Y}(x) \mathrm{d}x f(p) \mathrm{d}p \\ &= \int_m^{\infty} \int_{-\infty}^{p} f_{X|Y}(x) f(p) \mathrm{d}x \mathrm{d}p \\ &= \int_m^{\infty} \int_{-\infty}^{p} f_{XY}(x,p) \mathrm{d}x \mathrm{d}p\end{aligned}$$

其中 $f_{XY}(\cdots)$ 表示 X 和 Y 的联合密度函数。

现在，利用：

$$\int_m^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,p) \mathrm{d}x \mathrm{d}p = \frac{1}{2}$$

$\Big($由于内部的积分化简为 X 的边际密度，根据中位数的定义，外部的积分正好的等于$\frac{1}{2}\Big)$。那么，算法可以表示为：

$$\int_m^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,p) \mathrm{d}x \mathrm{d}p = \int_m^{\infty} \int_{-\infty}^{p} f_{XY}(x,p) \mathrm{d}x \mathrm{d}p + \int_m^{\infty} \int_{p}^{\infty} f_{XY}(x,p) \mathrm{d}x \mathrm{d}p$$

立刻，我们就可以得出：

$$\int_m^{\infty} \int_{-\infty}^{p} f_{XY}(x,p) \mathrm{d}x \mathrm{d}p = \frac{1}{2} - \int_m^{\infty} \int_{p}^{\infty} f_{XY}(x,p) \mathrm{d}x \mathrm{d}p$$

这看起来像是一个不相关的公式转换，但是实际上，只需两个步骤就可完成证明过程。此时，我们应用联合密度的对称性（对称性是由同边际分布的假设决定的）。在形式上，对称性的表达式为：

$$\int_m^{\infty} \int_m^{p} f_{XY}(x,p) \mathrm{d}x \mathrm{d}p = \int_m^{\infty} \int_m^{x} f_{XY}(x,p) \mathrm{d}p \mathrm{d}x$$

现在，调换积分的顺序（要小心观察积分的上下限），其代数式的等量公式为：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

$$\int_m^\infty \int_m^x f_{XY}(x,p)\,\mathrm{d}x\mathrm{d}p = \int_m^\infty \int_p^\infty f_{XY}(x,p)\,\mathrm{d}x\mathrm{d}p$$

因此：

$$\int_m^\infty \int_m^p f_{XY}(x,p)\,\mathrm{d}x\mathrm{d}p = \int_m^\infty \int_p^\infty f_{XY}(x,p)\,\mathrm{d}x\mathrm{d}p$$

倒数第二个步骤，注意后两个积分的和是1/4（再次根据中位数的定义得出）：

$$\int_m^\infty \int_m^p f_{XY}(x,p)\,\mathrm{d}x\mathrm{d}p + \int_m^\infty \int_p^\infty f_{XY}(x,p)\,\mathrm{d}x\mathrm{d}p = \int_m^\infty \int_m^\infty f_{XY}(x,p)\,\mathrm{d}x\mathrm{d}p = \frac{1}{4}$$

因此：

$$\int_m^\infty \int_{-\infty}^p f_{XY}(x,p)\,\mathrm{d}x\mathrm{d}p = \frac{1}{2} - \int_m^\infty \int_p^\infty f_{XY}(x,p)\,\mathrm{d}x\mathrm{d}p = \frac{1}{2} - \frac{1}{2}\left(\frac{1}{4}\right) = \frac{3}{8}$$

第二个部分的推导过程也是类似的。

4.4.2 示例

[示例1] 假设序列的相关性参数 $r=0.71$（见图4-3，关于这一假设见4.4节中相关论述），以及一个正态分布随机项，从一阶自动回归模型中抽取1000个随机样本。图4-4b显示了时间维度的图形；图4-4a显示了样本的边际分布。

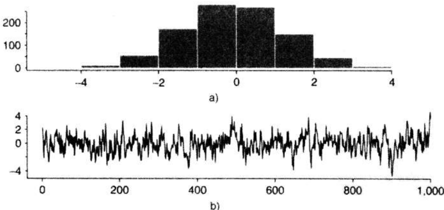


图4-4 从自相关模型抽取的随机样本

这个序列的反转移动比例为62%。

更加实际一些，我们把这个序列当成是每天的观察值，计算出一个估计中

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

位数。用局部的中位数调整数据序列（使用的周期为10)，对序列进行分析，如图4-5所示。这个调整过的数据序列呈现一个较大的反转移动比例，为65%。

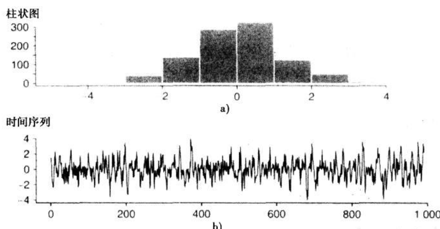


图4-5 从自相关模型抽取的随机样本，根据局部中位数调整过的图形

4.5 非常数分布

假设100天的价差序列是由一个正态分布所产生的，此后50天价差序列是由一个均匀分布所产生的。在各个期间，基本的定理都适用；因此，在这150天中，除了有一个例外的情况之外，75%规则是正确的。假设在第151天，价格范围又由正态分布来产生。在此情况下，我们该怎么办？

我们不含糊地说，就平均水平而言，75%规则在整个序列期间是正确的。在两个转换的日子里，这一规则不成立：①从正态分布变成均匀分布的那一天；②从均匀分布变成正态分布的那一天。回顾图4-1。对于根据任何两个不相同的、连续的和独立的分布而言，产生出来的随机量在区域1的概率上，会被 $0<\Pr[(1)]<\frac{1}{4}$ 所限制（意味着 $\Pr[(1)]=\Pr[(P_t,P_{t-1})\in(1)]$）。这里遵循了在第4.2节所描述的中位数定义。用 p 来表示这个概率。现在，在象限中区域1的补集概率是 $\frac{1}{4}-p$（与前面的“区域概率”的

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

意义相同)。转换是关键问题，因为只有分布变化会改变基本结果。当然，证明规则不成立也是很容易的。当在转换的时点 $p > \frac{1}{8}$ 的情况（从正态分布转换到均匀分布的时点)，定理的结论就不会是 75%，而是 100 $(1-2p)\% < 75\%$。然而，对于每一个转换，在区域 1 中，反向的转换会出现补集关系的概率，也就是$\frac{1}{4}-p$。

类似的分析可以在区域 2 中加以应用。不管区域 2 的概率与区域 1 的概率是否相同，只要两个分布其中有一个是非对称的分布，这个结论都是正确的。

因此，如果转换以成对的方式发生，表现出来的概率是平均的，75% 这一结果会重复出现。(如果两个密度函数都是对称的，那么“成对出现”这个条件就足够了。但是，如果有一个密度函数是非对称的，$\Pr[(1)] \neq \Pr[(2)]$，那么“成对出现”这个关键的条件就必须经常发生，在统计上保证“成对的转换点”，落在区域 1 和区域 2 的概率小于落在区域1－1或 2－2 的概率)。

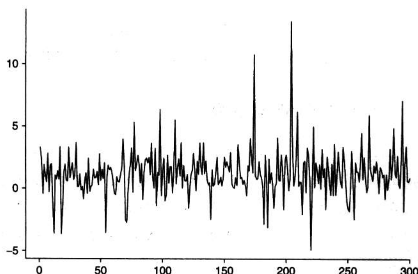


图 4-6 从正态、对数和学生 t 的分布中抽取的随机样本

将这个结论更进一步扩展，在三种各自不同分布的情况下证明这个结

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

论。关键的条件是，在三个分布中，两两转换发生的概率是相同的。对于任意数量的不同分布，应用数学归纳的方法，可以证明理论是正确的。

4.6 结论的应用

在数页的理论推导之后，暂停一下，问一个问题："这些结论与以模型为基础的股票交易有什么关联?"之前，一个主要的假设是关于稳定性假设——到目前为止，为了便于讨论，这一假设没有提及。我们要求价差必须是"独立的，同分布的"（后来放宽了这个条件，数据序列可以是相关的）；其实暗示了一个假设，那就是在时间序列中，存在稳定性假设。

股价序列并不是稳定的。那么价差呢? 一般来说，价差也是不稳定的。但是，对于局部位置和方差的估计值进行过动态的调整，可以得到一个合理的、稳定的近似值。这与协整（见第3章）的观点之间有联系。揭开单个价格序列中的分布结构，可能是一件很难的事，但是预测序列之间的差额（价差）是比较容易的。扩展协整的基本概念到局部定义的序列（我们可以使用"局部协整"这一术语），我们能够识别出其中的联系。

应用动态变化的近似值这个原则，理论上的结论在指导性上具有有效性。

4.7 应用于美国债券期货

尽管本章讨论的是相似公司的股票价差（传统的匹配交易方式），但所展现的定理能被广泛地应用。只要"价格"序列的条件相符，这个定理能应用到任何金融商品之上。当然，关键点在条件的符合上——有很多的序列不符合（在时间上，没有显著的结构性变化，例如趋势变化）。债券的价格符合这个定理的条件。

利用简单的预测模型，研究美国30年期公债期货每天的价格变化的范围。观察当前最具有流动性的期货合约。（由于考虑到合约到期时，可能会发生价格扭曲的现象，在期货到期前15个交易日，利用展期的方式扩展合

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

约的时间。在分析中，就不会观察到扭曲的现象，因此采用这个序列进行展示。）图4-7显示了1990～1994年研究数据的样本分布——在这个分布中有一个明显的斜率。这个序列在时间上所展现的图形如图4-8所示。

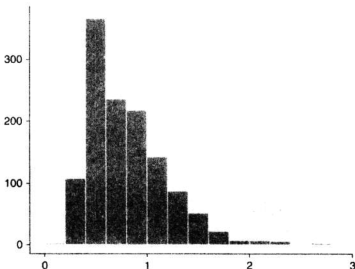


图4-7 美国30年债券期货的价格变化范围的边际分布

在这个预测模型中，中位数是根据前20个交易日的数据计算出来的。在操作中，计算出的局部中位数变化相当小，这颇为令人满意。因此，一个优点是序列的相关性较低：原始的序列（等同于使用一个常数作为中位数）在[0.15，0.2]的范围内，显现出自相关性；但如果是用局部中位数调整后的序列，则呈现出显著的自相关性。

预测执行的结果展现在表4-1中：预测的结果与定理可以说是一致的。

表4-1 美国30年债券的经验研究 (%)

年份	比例 $P_t > P_{t-1} \mid P_{t-1} < m$	比例 $P_t < P_{t-1} \mid P_{t-1} > m$	总比例
1990	70	75	73
1991	77	72	74
1992	78	76	77
1993	78	76	77
1994	78	76	77
全部	77	75	76

注：每年有250个交易日。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

债券期货价格的结论，在经济上有利用价值。在应用这个交易规则时，需要考虑一些精密而复杂的前提条件，必须要特别注意的是管理交易的成本，但这有许多的可能性。

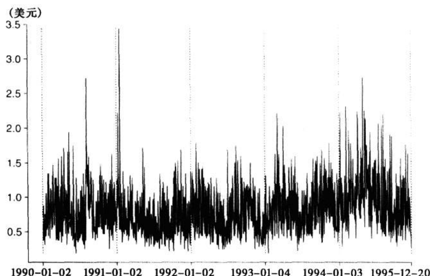


图 4-8 美国 30 年债券期货每日的变化范围值

4.8 总结

这个定理的含义具有很强的煽动性：只要符合这个定理的条件，保证有 75% 的预测精确度。在实践中，在很短的期间内，很多的股价和价差会符合上述条件。此外，当变化发生时，变化发生的概率必须足够的慢，导致校验后的动态模型（在本章中，所展示的例子是对于均值的局部性描述）能出现理论所预测的反转结果。

根据 30 年期美国债券上的经验证据，暗示这个市场能接近于符合条件的情况。1990 ~ 1994 这五年的经验模型，与 75% 之间，没有不同的地方。只要应用一些精巧的转换，很多看起来违反这个定理条件的情况，都可以被转变成近似满足条件的情况。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

附录 4A：向前预测几天

假设价差没有持续向某个方向移动（趋势），正如我们在理论发展过程中所做的那样，在应用中对局部位置的数据进行调整，如果向前考虑一天以上的时间，就存在比较大的反转机会。当然，很重要的是，在 k 天的期间内，每天的情况都可以单独地观测到；如果预测 k 天之后的移动情况，那么就跟预测一天的情况一样，在本质上没有变化。如果考虑每天的情况，就会有很多（独立的）赢的机会（如一次反转），所以赢的概率就会增加。

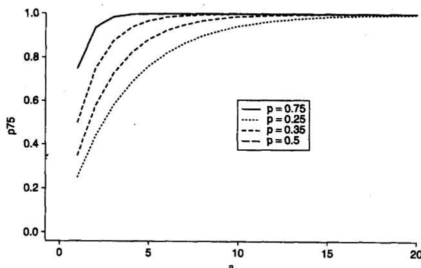


图 4-9 在 n 天中，至少有 1 次“向赢方向变化”的概率

图 4-9 展示了在 k 天中，从目前的位置往中位数移动的概率，其中 $k=1$，…，20。根据定理，在一天之内，发生这样的一个移动的概率是 0.75。其他的概率，可以用稍后介绍的二项式分布来计算。图中也包括了赢的概率分别为 0.5、0.35 和 0.5 时的图形：在实践上，最好的情况下观测到的数值会低于 0.75，这是因为无法完全地符合定理的假设条件。

假设每天的价格是独立的。对于任何一个时点，这个定理认为，从目前的位置发生反转的概率都是 0.75。因此，对于 $t+1$ 时点，t 时点的价格开始出现反转的概率是 0.75，同样的，不管 $t+1$ 这天发生了什么事（这就是独

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

立性的假设)，对于 $t+2$ 时点也是相同的，诸如此类。因而，在 k 天中，有几天的价格，会相对于今天的价格显示出反转，这个结果服从二项式分布，具有参数 k（试验的次数）和 0.75（一个给定的试验成功的概率）这两个参数。显示在图中的概率计算公式如下：

$$\Pr[\text{在 } k \text{ 天之中，成功超过 1 次以上}] = 1 - \Pr[\text{在 } k \text{ 天之中，成功 0 次}]$$

$$= 1 - \binom{k}{0} 0.75^0 (1 - 0.75)^{k-0} = 1 - 0.25^k$$

如果放宽独立性约束条件，二项式的结论就是不正确的，所产生的偏离度是序列时间结构的一个函数。结构函数是一个简单的趋势，那么往前预测的天数越多，二项式计算的概率的准确性就越低。对预测函数的结构形式予以适当关注，例如加入一个估计的趋势成分，可以提高预测的准确度。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

第 5 章 高斯不是反转之神

大致正确比完全错误要好。

——凯恩斯

5.1 导论

我们从两段引用的文字开始：

名义利率（nominal interest rates）的分布，显示利率并没有向平均值反转（平均回归），其分布的结构不是正态分布。

与之相对应的是：实际利率（real interest rates）呈现出正态的分布迹象。这样的分布说明了利率具有向平均值反转的特性。

这两段引用的文字，出自于某个大型而且成功的华尔街公司几年前公布的一份研究报告，存在逻辑上的错误暗示。由于没有明确定义“均值反转”的含义，因此有人可能会把收益的边际分布当成是反转的定义。但是，这意味着作者没有用一种广泛的讲解方式，来讨论“均值反转”。

第一个描述是错误的，因为无论时间序列的边际分布呈现出什么形状，都可以是均值回归。第 4 章中的 75% 规则很明白地证明了这一现象。请注意“可以”这个词语。声称一个时间序列具有均值回归的特性，仅考虑时间序列的边际分布是不够的。这也就是第二个描述错误的原因。某个时间序列与

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

想象中均值回归的现象（完美的回归现象）相差得很远，而这个时间序列却又服从正态的边际分布，这是一种完全合乎情理的现象。在本章的最后一节，以举例的方式对此进行说明。

按照均值回归的定义，它随着时间呈动态变化：从边际分布中抽取的时间序列样本，其抽取的顺序是关键的因素。

5.2 双峰骆驼与单峰骆驼

本章开篇所提及的研究报告，说明了一个从 1953 年开始，10 年期债券收益的柱状图，其收益出现在一个名为“双峰分布”的小节中。其研究认为，非正态边际分布的时间序列不会出现反转的现象。不管“双峰分布”的比喻意义，事实上，柱状图存在着好几种模型。不过，拿它来作为例子，说明服从双峰分布的时间序列具有反转的特性，论据是很充足的。

图 5-1 显示一个样本，这个样本是从均值为 4.5 的正态分布中随机抽取 500 个样本，然后从均值为 8.5 的正态分布中随机取样 500 个样本值（两个分布的标准差都为 1），组合后的样本。在观察债券收益分布的主要模式之后，提出这些前提条件。对于本章的示例来说，不是很重要的事情。

图 5-2 显示了这 1 000 个点可能出现的一个时间序列。从第一天开始，这个序列出现了多少次反转？事实是出现了很多次反转。我们将分别考虑每个部分。在任一部分中，分别取得远离均值的点（第一个部分的均值是 4.5，第二个部分的均值是 8.5）。时间序列中会出现一个比较接近均值的点。那么这一过程要经过多少个点？一般来说，要经过 1 ~2 个点。这个序列从来没有发生过，远离均值之后，不再向均值回归的现象；在相反的情况，这个序列反复地在均值附近震荡——这可以说就是一个反转的定义。这正是一个真正的、令人兴奋的爆米花过程！

剩下需要考虑的事情是那个单一变动点的含义：序列从第一个部分移动到第二个部分的那个点，到第二个部分均值就增加了。在一个时间序列中，存在一个均值平移，就可以认定其不具备均值回归的特性吗？这个时间序列

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

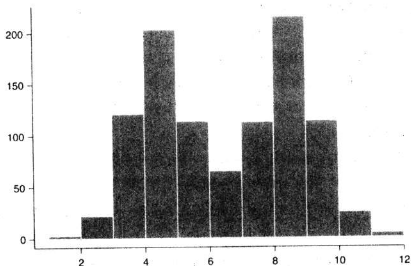


图 5-1　边际分布：混合 N［4.5,1］和 N［8.5,1］两个分布之后的情形

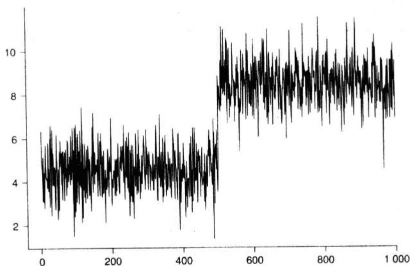


图 5-2　在图 5-1 中所描述的样本呈现出来的一个时间序列

没有回归到整体平均值——但对于均值回归的含义进行限定是不寻常的；它对于理解一个时间序列上下关系来说是没有帮助的，而且容易产生误导。当我们在评估一个时间序列中的交易机会时，这种做法也会产生误导。假设原

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

始数据是每天的数据。在两年的时间里，第一个部分的数据肯定会被描述为具有均值回归的现象。平均值突然增加，但很快就会得到相同的结论（出现均值回归现象），虽然序列回归的均值变大了。在两年多的时间里，反转到（新）平均值的现象重复出现。如果每日数据服从边际分布，按照基本的均值回归策略进行交易，并获得了丰厚的回报，谁会否认这个序列具有均值回归的现象呢？

第二个分布的关键点是，债券数据分布的右边呈现“厚尾”现象。实际上，这个特征与均值回归的特性是不相关的。图 5-3 显示了有 1 200 个点的柱状图：其中 1 000 个点来自前面的正态混合分布，然后再加上 200 个服从区间为［10，14］的均匀分布的随机数据。图 5-4 显示了这 1 200 个点可能形成的一个时间序列：在这个序列中，有多少反转？有很多反转。在前面的段落中已经论证过，前两个部分具有均值回归的现象。那么最后的那个部分呢？这些点是来自于一个均匀分布的随机样本，而不是一个正态分布，因此根据最前面的研究报告，这个时间序列不具有均值回归的现象。如果你同意这种观点，我愿意同你赌一把。

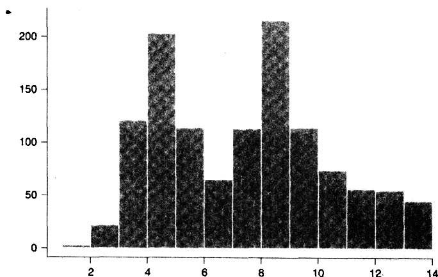


图 5-3　右边具有“厚尾”现象的边际分布

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

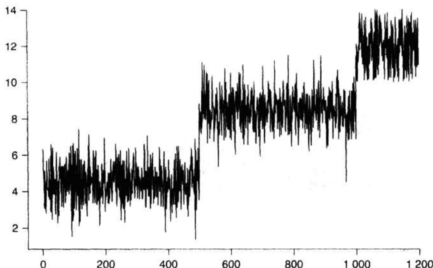


图 5-4 在图 5-3 所描述的情况下，按照样本分布的顺序，抽取样本所形成的时间序列

当然可以认为这些示例都是人造的，不真实。真实的时间序列不会出现如此便于识别的图形。这是无可否认的。但这也不是一个问题。图 5-5 显示了按照图 5-3 的分布，随机抽取样本形成的另一个可能的现象。它完全是随机的选择，没有做人为的替换工作，而这 1 200 个原始的样本，按照选择的顺序进行绘图。有多少的反转现象？存在很多反转。可以用你的眼睛来分析判断，你会不同意吗？

时间序列的边际分布是单峰还是双峰，不影响判断它是不是具有均值回归的现象。重复一次：时间的结构是关键的特性。

干涸的河流

骆驼与其他动物不同之处，在于能适应干旱地区的生活。它有两个重要的能力：一是经过进化、像海绵一样吸收水分的身体，二是能慢慢地输送养分的能力。干涸河流在一段时间没有任何水流，这种情况在干旱的地区十分常见（干涸的河流经常是骆驼补充水分的来源）。长期观察一条干涸河流，会发现河流的水位所形成的时间序列具有反转的现象。不管在雨季时水位偏

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

离平均值有多远，时间序列的值还是会回到零。

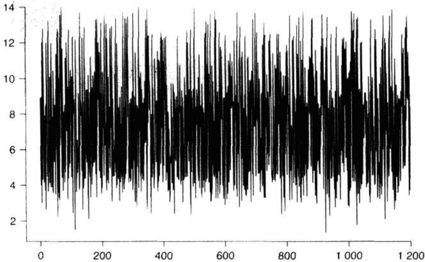


图 5-5 在图 5-3 中所描述的情况下，随机抽取样本所形成的时间序列

干涸河流的典型边际分布是什么？显然它是非常不对称的。所以，在完全没有借助数学公式的情况下，我们证明了一个具有反转特性的时间序列，并不需要具有像正态分布那样对称的边际分布。

丧钟还在敲响。通常用一个所谓的伽马分布作为间歇性河流的数学表达式（允许合理地取值为零）。现在，将高斯变量取平方，其分布就变成一个卡方随机变量，同时也是一个伽马随机变量。还有更令人好奇的巧合吗？

但丧钟不是为反转而响：干涸河流的示例，只是一个特殊的案例，能自我进行证明存在反转现象。重新声明一次：任何边际分布都存在具有反转特性的时间序列。

5.3 依然在敲响钟声

一个展现出正态边际分布的时间序列，不一定具有均值回归现象。特殊的转换也是可能的。图 5-6 显示了从正态分布中抽取的随机样本，正态分布

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

的均值为 2.65，标准差为 2.44——在引用的数据中，这些样本的数值是真实的收益数据。图 5-7 显示了这个样本可能形成的一个时间序列，样本点是按照大小进行排列。从第一天开始，在时间序列中，出现多少次的反转？一次都没有。与先前的示例对比，可以指责这并“不实际”，或者“图形中的点不会按照大小的顺序出现”。但是这样的指责是无关紧要的：这个示例说明一个事实，不管从分布的图形中得到什么，即便时间序列样本具有正态的边际分布，也无法揭示该序列的时间特性，比如反转现象，或者任何其他特性。

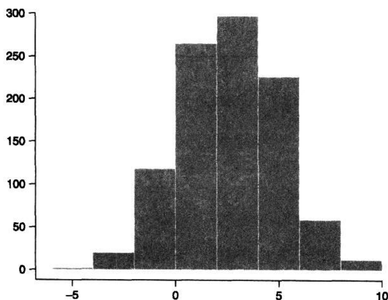


图 5-6 从 N [2.65, 2.44^2] 中抽取 1 000 个随机样本

本章开篇引用的论文，以及在本章中进行批判性检验的内容，没有出现在本书的参考书目中。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

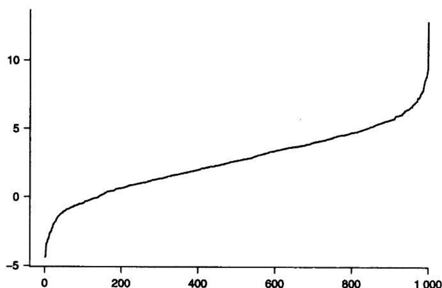


图 5-7 在图 5-6 中所描述的情况下，所形成的时间序列

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net



第6章 价差波动率

投资者会（或者应该）将回报的期望值视为一个想要的东西，而把回报的方差视为一个不想要的东西。

——诺贝尔经济学奖获得者 哈里·马科维茨

6.1 导论

匹配交易中的反转现象，是指两只股票的价格分开之后靠拢，或者是两只股票的价格非常靠近之后再分开的现象——例如第2章中谈到的爆米花过程。股价移动的数量由股票的波动率进行计量，波动率是指价格变化（回报）的标准差。跟价差反转方案有关的波动率，就是股票之间相对价格变动的波动率，简称为价差波动率。图6-1a显示了在1999年上半年，两个相关的股票（ENL和RUK），每天的收盘价格（这是已经按照股票分割与股息的状况，调整过的价格）。这两只股票价格的变动密切相关，在图6-1b中的价格差额（价差）序列中，可以进一步说明。忽略数值范围上的差异，图6-1b中的价差序列看起来像是任意的股价序列。价差的波动率在特性上与股价波动率极为相似，这一发现并不令人感到意外。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

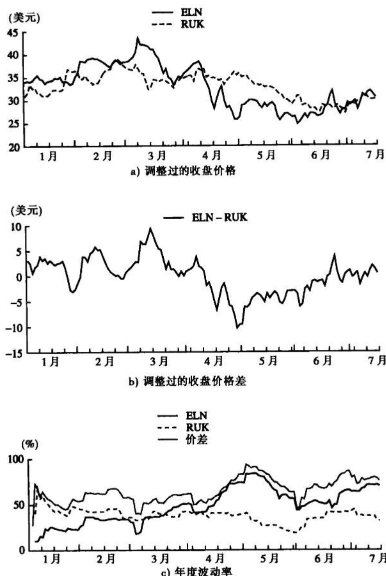


图 6-1 ELN 和 RUK 股价走势（1999 年）

股价的波动率就是回报的标准差。价差波动率的计量标准是什么？在第2章的时候，我们直接用价差自身的标准差校验交易规则。在这里，我们感兴趣的是波动率在技术上的定义，不仅仅是当成一个衡量的因素，需要把焦点放在一个合适的回报序列上。考虑图 6-1b 中的价格差额，这个差额围绕

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

着零点上下波动。显然我们不应该将价差当成一个价格——这样一个序列很容易出现无限“回报”的情况。利用价差，找出一个计量基本的反转策略的方法：当价差大于或小于某个极限值时，预测会有反转现象，此时进行两笔交易，每一只股票进行一笔交易，其中买进一只股票，卖出另外一只股票。这样，价差交易的回报就是买进股票的回报减去卖出股票的回报：

价差回报 = 买进股票的回报 − 卖出股票的回报

（假设两笔交易的金额相同，并且只计算多头的回报）。因此，价差交易值变动的范围（或者是价差的波动率）的计量方法，可以直接根据价差回报序列计算出来，也就是买进与卖出股票回报的差额（深入到细节之中，我们可以看出，如何将这个针对匹配交易的方法，推广到其他的统计套利交易中）。

图6-1c显示采用20天的时间周期，ENI和RUK这两只股票，以及ENI-RUK价差的波动率轨迹。（在本章所有的例子中，计算波动率都是依照惯例，假设零为局部的平均回报。）价差的波动率，像预先给予的暗示一样，在图形上与股票的波动率十分相似。令人奇怪的是，价差波动率一直都大于两个股票的波动率——尤其到了后面。

在图6-2中，显示另外一个例子，这次将通用汽车（GM）和福特汽车（F）作为匹配交易的对象。请注意，价差的波动率有时候比两只股票的波动率大一些，有时候又比两个股票的波动率要小一些，有时候又介于两个股票波动率之间。

这两个例子揭露了价差波动率的很多特性，不管针对最简单的匹配交易这样的示例，或者是一篮子股票之类的示例，这些特性对于理解和利用价差关系，都是很重要的。图6-3显示了另外的一个例子，这次选择的是两个不相关的股票，分别是微软和埃克森。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

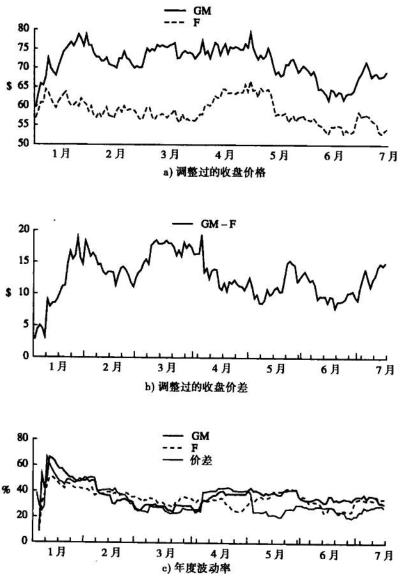


图 6-2 通用汽车与福特公司的股价走势（1999 年）

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

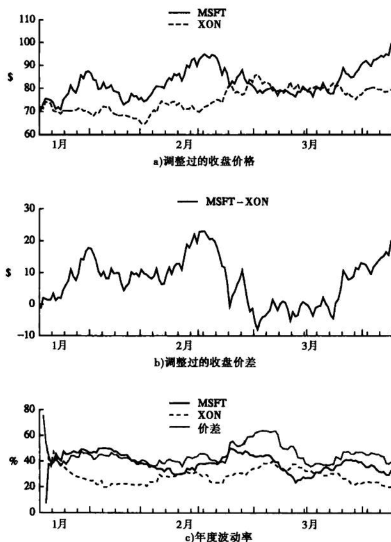


图 6-3 微软和埃克森股价走势（1999 年）

6.2 理论上的解释

股价之间的相对变动与个别股票的价格变动之间的依赖关系，可以用函数来表达。这样的话，不是比"A 的价格 - B 的价格"更简单吗？事实上，如果观察股价之间的相对变动的波动率，就会发现它们的关系

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

更加错综复杂。因此，在价差回报中，最关键的公式在前面已经给出来了：

价差回报 = 买进股票的回报 - 卖出股票的回报

用 A 代表买进股票 A 的回报，B 代表卖出股票 B 的回报，S 代表价差回报，价差波动率可以符号描述如下：

$$\sqrt{V[S]} = \sqrt{V[A-B]} = \sqrt{V[A] + V[B] - 2V[A,B]}$$

其中符号 V[·]表示（统计学或概率论上的）方差，而符号 V[· , ·]表示协方差。这个数学公式揭示，价差波动率分别小于、大于或等于个别股票波动率的理由和原因。一个关键的因素就是两只股票（回报）的协方差，即 V［A，B］。

如果在现实中，A 与 B 这两只股票（故意滥用这两个符号）是相同的股票，那么两个股票的方差（波动率的平方）就是相同的，而且至关重要的是，协方差等于单个股票的方差。因此，价差的波动率就会等于零：由于同一股票的价差本来就等于零，还会有别的原因吗？

如果这两个股票是完全不相关的，会出现什么情况？从统计学的观点来看，这等于是说协方差为零。那么，价差波动率的公式就可以化简为：

$$\sqrt{V[S]} = \sqrt{V[A] + V[B]}$$

这也就是，价差的波动率会大于两只股票波动率。如果单个股票都具有相似的波动率，也就是 V［A］ = V［B］，那么价差的波动率会增大 40% 左右：

$$\sqrt{V[S]} = \sqrt{V[A] + V[B]} \approx \sqrt{2V[A]} = 1.414\sqrt{V[A]}$$

6.2.1 理论与实践对比

在图 6-1 中的示例显示了，两个“相关”股票之间的价差波动率大于单个股票波动率的情况。前一节中的理论认为：①对于两个完全相同的股票，其价差波动率等于零；②对于完全不相关的股票，其价差波动率将会大于两个股票的波动率。可以肯定，“具有相关性的股票”（如 ENL 与 RUK）的价

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

差波动率，不像“完全不相关的股票”的价差波动率，而更像“完全相同的股票”的价差波动率。根据上面的理论，ENI-RUK 价差的波动率不应该要小一些吗？

现在，我们必须对一些术语，将其统计上的定义和字面上的含义进行区分。艾斯维尔（Elsevier）集团的两只股票，也就是 ENI 与 RUK，确实是相关的——本质上他们是同一家公司。两个股票的价格序列的历史轨迹非常地相似。在很长的一段时间里，可以说它们两者的价格是相同的。然而，以每天为时间单位进行观察，两个股票的价格轨迹很少同步移动；因此两者之间的价差确实是变化的（见图 6-1b），而且价差的波动率也为零。事实上，从短期来看，这两个价格序列呈负相关性：在图 6-1a，这两个股票的价格轨迹，就好像动画片中的两条相互缠绕的蛇，当一个往下移动，另一个就往上移动。从统计学上来说，在计算局部波动率时，尤其是在衡量短期回报的尺度上，这意味着两个序列是负相关的。

负相关性（也就是，协方差为负数），把它放到价差波动率的公式中，立刻就能很清楚地理解，为什么艾斯维尔集团的两只股票价差波动率会大于单个股票波动率了。利用匹配交易，可以从中获得收益！

6.2.2 进一步完善这个理论

回到价差波动率的数学表示式：

$$\sqrt{V[S]} = \sqrt{V[A] + V[B] - 2V[A,B]}$$

假设 $\underline{\sigma}^2 = \min(V[A], V[B])$，$\underline{\sigma}^2 = \max(V[A], V[B])$，对于不相关的股票（$V[A, B] = 0$）来说，其价差波动率会处在单个股票波动率乘上某个数值之间：

$$\sqrt{2}\underline{\sigma} \leqslant \sqrt{V[S]} \leqslant \sqrt{2}\bar{\sigma}$$

在前面，谈到两个观察结果：对于两个波动率很接近的股票，其价差的波动率会比个别股票的波动率大 40% 左右；对于两个完美正相关的股票，其价差会是一个常数，所以价差的波动率等于零。还有一个与相关性有关的极端示例：当 A 与 B 是完全“负相关”，也就是 $A = -B$，A 与 B 的价差波动

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

率是单个股票波动率的四倍：

$$\begin{aligned} V[S] &= V[A] + V[B] - 2V[A,B] = V[A] + V[-A] - 2V[A,-A] \\ &= V[A] + V[A] + 2V[A,A] = 4V[A] \end{aligned}$$

因此，$\sqrt{V[S]} = 2\sqrt{V[A]}$。对于统计套利来说，这（几乎）就是价差关系中的“圣杯”。

6.2.3 应用到示例中

从 GM-F 和 MSFT-XOM 这两个例子中，应用价差波动率理论的优点还能推导出一些别的什么吗？假如描绘股票与价差波动率的轨迹，我们可以在价差的变化之中，指出相关性为正数和负数的区间。当然，可以用直接计算的方式，观察相关性（见图 6-4）。在 1999 年的上半年，相关性的平均值是 0.58；针对 20 天周期相关性的最大值为 0.86；针对 20 天周期相关性的最小值为 0.15。这里要注意的是，对于运用价差来说，相关性、价差波动率和方差范围的动态变化，都是非常重要的。

从 4 月下旬开始，GM-F 的价差波动率就一直低于单只股票的波动率，与 3 月大部分的时间一样。事实上，从图 6-2b 中，在 4 月、5 月、6 月大部分的时间里，同其他期间比较起来的话，价差的轨迹是一个常数。但是在图 6-2c 中，价差波动率轨迹在 4 月份出现了一个 50% 左右的增幅，在 5 月份出现了一个类似降幅。4 月 7 日那天（见图 6-5），出现了一个特别大的（负）单日价差回报，加上采用 20 天的周期计算波动率的估计值——回顾图 6-4 的局部相关性，就明白这是一个人为的结果。对于波动率这种间接的衡量指标，会使用外部权重剔除法和平滑法等典型的方法，来平滑这种不连续跳动的曲线（见第 3 章）。图 6-5 显示的是价差回报：也就是 GM 的回报减去 F 的回报。

6.2.4 衡量价差波动率的基础知识

让我们先问一个问题：当波动率较高或较低时，统计套利能产生出较高的回报吗？

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

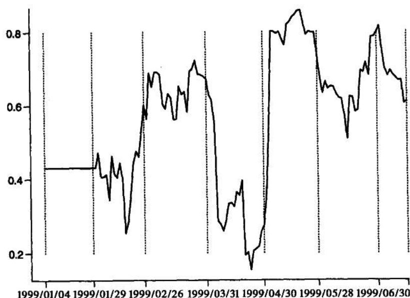


图 6-4 以 20 天为时间周期，滚动计算 GM-F 的相关性

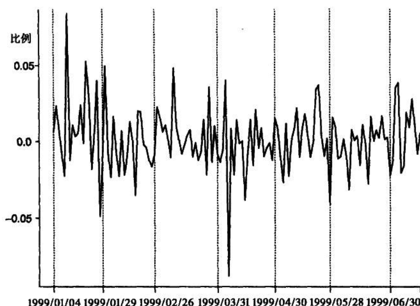


图 6-5 GM-F 价差的每日回报

不考虑与股票相关的任何特殊事件，而且根据一个经过很好校验的模

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

型，比较高的价差波动率会产生出比较大的回报。图 6-6 显示了 1995～2003 年，成对股票价差的平均局部波动率（取 20 天的移动周期）。这些成对股票是从标准普尔 P500 指数的股票池中选择出来的。对于统计套利来说，2000 年与 2001 年是回报最好的两年。这两年的价差波动率都很高；2000 年的价差波动率和统计套利的回报，比 2001 年要高——这似乎支持了前面的说法。1999 年的价差波动率也是很高，但当年的统计套利回报，是 10 年中最差的一年。在 1999 年，有很多与股票相关的特殊事件，这些事件主要与收益相关，都对回报产生了负面的影响。这些事件是如此的引人注目、普遍和影响恶劣，以至于美国证券交易委员会，最终通过了所谓的公平披露规则，宣布这些行为是不合法的。

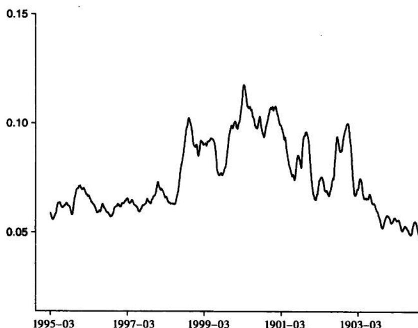


图 6-6 局部价差的平均标准差

如果使用波动率的局部估值，那么会从具有代表性的价差序列上得到什么图形呢？利用样本数据涉及的波动率模式，能从图 6-6 的价差波动图形中做出什么推断呢？

图 6-7 说明了两个价差样本序列的局部波动率（使用相同的权重，采用

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

20天的周期）。最上面的图6-7a展示的是价差序列，而中间的图6-7b是局部波动率的估值。这里并没有什么特别令人惊讶的地方，从最上面的图形中，取一段固定长度的曲线，然后计算变动的衡量结果。请注意，我们观察到，这两个序列局部波动率的平均值是很接近的。

当使用不同的方式衡量“局部值”，会发生什么情况呢？图6-7c最下面的图形说明了采用60天周期的状况：现在存在一个惹人注目的特征，波动率比较高意味着价差的振幅比较大（虽然我们一直都是采用价差进行描述相关的陈述，但这里的讨论可以应用到任何时间序列之中）。同样的，这里并没有什么特别令人惊讶的地方。采用60天作为时间周期，截取了一个完整的循环，这个循环具有较大的振幅序列——如果截取的是一个完整的循环，其波动率的估值将会是一个常数。因此，估计的局部波动率反映出了时间序列的振幅大小。在之前的示例中，比较短的时间周期仅仅反映了移动比较慢的序列的变动趋势。哪一个估计的波动率能反映反转回报的机会呢？在这里，答案是很简单的。

如果检查这两组序列的平均值，而每一组序列跟自身的“噪声”在振幅和频率上相混合，这样的话，会出现怎样的景象呢？

提醒一下，从图6-6中能推断出什么呢？在尝试给出一个答案之前，原型示例的分析，清楚地建议，从某个时间周期的范围（或某个局部加权方案），注意局部估计的波动率——将证据集中在“比较短的区间”，并将焦点放在局部波动率的平均值上，这看起来是一种比较明智的做法；估值中普通的波动“可能”与人为的扰动没有什么不同。

图6-8再次复制了图6-7中两个示例的价差曲线，并增加了第三个示例。新序列的振幅与原来振幅较大的序列相同，而新序列的频率与原来频率较高的序列一样。因此，新序列有比较高的频率，反转机会的价值较高。中间的图6-7b，描述局部估计的波动率，指出了第三个序列的平均波动率，正如我们所期望的，是原先两个波动率的两倍。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

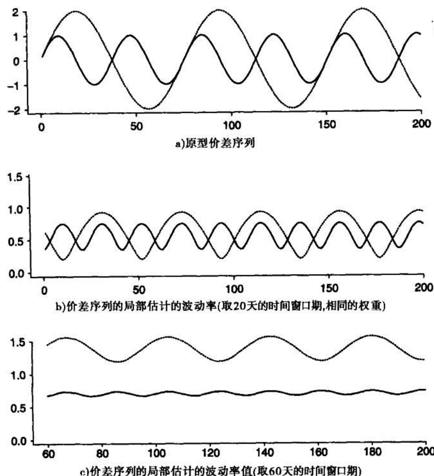


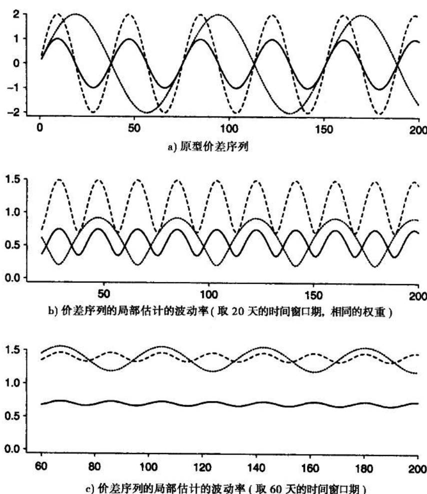
图 6-7 价差曲线示例一

现在看一看最下面的图 6-7c，它显示了采用更长的时间周期，所得到的局部估计的波动率。有趣吗？这个分析再一次指出，当根据价差波动率的平均水平，来推断反转的机会，应该使用更短的时间周期、更狭窄的视野。

应用这些原型，我们可以采用恰当的“时间—频率”分析方法，对反转移动的数量进行精确的量化。现实中的价差序列不会那么友好，通常伴随不稳定的状况和噪声的“污染”。

前面的评论只是一般性描述，说明了序列的变动是如何通过经验反映在统计摘要上，并指出简单反转中潜在获利的大小是如何估计出来的。无论是

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net



a) 原型价差序列

b) 价差序列的局部估计的波动率（取 20 天的时间窗口期，相同的权重）

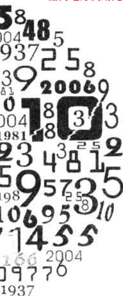
c) 价差序列的局部估计的波动率（取 60 天的时间窗口期）

图 6-8　价差曲线示例二

在此使用的理想化的无噪声正弦序列中，还是在应用于真实的价差历史数据上，我们在实际运用反转策略时，都必须对策略进行直接的分析，才能得出一个合理的推论。

以自 2002 年以来统计套利的表现每况愈下为背景，在第 9 章中我们还会再次讨论价差波动率。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net



第 7 章 将反转机会量化

我们用持久的忍耐力征服一切。

——极地探险家　沙克尔顿家族座右铭

7.1　导论

本章，我们对前几章所涉及的理论进行进一步的拓展，以便能够更深入地掌握时间序列中反转的特点。与本书其他章节相比，本章会涉及更多抽象的概念，以及更高深的数学知识。本章所有的示例，及其理论上的发展，最终都运用到实践中去。大多数的讨论，都是在价格序列的基础上进行的，不过，这些讨论的结果，一般也可以应用到其他的时间序列中去。特别值得注意的是，有时候只要对某些推论稍作改动，就可以轻而易举地将此分析应用到价格序列的转换之中，包括回报、价差组合、价差回报，以及因子分析，等等。

关于“什么是反转”这个问题的讨论，在此之前我们都是假设一个概率模型，并产生股价数据来进行说明的。这个模型已经被高度程式化并过度简化，所以只能被当做是阐释反转概念的一个工具。我们完全可以利用模型得到一些数据，来估计真实而未知的价格，其实这并不是空想。假定我们对这些简单模型严格限定，并重新检查反转的含义，这时我们可以对“反转是什

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

么”这个问题看得更清楚。我们希望能从这样一个有意义的检查中，对可计量的反转概念提出一些有用的概念，运用在“价格生成模型”的相关研究之中。希望这样的分析，可以为我们提供出一些关于统计套利的新见解，协助我们从整个系统或运作机理等方面，进一步的分析并了解统计套利运作的方式和原理。然后以这些见解为基础，为驱动这些机会的力量，建立起一个有效的市场理论。这么说有点雄心勃勃，或许是希望能找到一个理论或过程，可以带来超越这些指标的作用。如果我们想要研究并解决从 2004 年起就一直困扰着统计套利的问题，并了解目前还影响着绩效表现的原因，那么了解机制和理论都是相当重要的：市场结构的变化是如何影响绩效表现?

7.2 平稳随机过程中的反转现象

我们先研究最简单的随机过程，也就是平稳随机过程。假设每天的价格，都是根据相同的概率分布，独立生成，这个分布的特性是由固定的参数决定的。先假设它是一个连续的分布。而 P_t 代表着 t 这天的价格。这里采用大写字母 P，是因为小写的字体要表示一些特别的值（例如实际的价格）。

这样假设之后，我们马上就可以得出以下几点：

（1）如果 P_t 处于分布的尾部，那么 P_{t+1} 就可能比 P_t 更接近分布的中心。用正式术语说就是：假设 $P_t > 95\%$。那么 P_{t+1} 小于 P_t 的概率就是 95:5 (19:1)。落在不同的百分比位置，也可以得到相似的结论。

“P_{t+1} 有 19:1 的概率小于 P_t” 与最前面的 “P_{t+1} 可能比 P_t 更接近分布的中心”，这两种说法，它们的含义并不完全相同。“更接近中心”与“小于”是不同的。某些概率以及其隐含的意思的观点有时候是很有意思的。从完整性的角度看，检验是否“更接近中心”是很有用的。“更接近中心”是比较正式的：它可以以价格为标准，衡量偏离中心的大小。衡量偏差大小，也可以采用另外的方案，用价格分布百分比来表示。这两种概念对于对称的密度分布函数来说是一样的：不过对于不对称的函数，就大相径庭了。

（2）如果 P_t 很靠近分布的中心，那么 P_{t+1} 就很有可能会比 P_t 还要偏离

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

分布的中心。

再仔细地想一下，就会发现，第 2 点对于一个反转研究来说并没有什么作用；但如果按照文章的思路来说，接近中心的值，还是可以标识偏离中心的情况，也即是标识未来的反转机会，因此它还是有用的。回顾第 2 章的爆米花过程和第 3 章随机共振的知识，你应该有更深刻的体会。

将第 1 点中的概念进行推广，可以为本章问题的研究提供一个潜在的起点：假如 $P_t >$ 分布的 $p\%$，那么 $P_{t+1} < P_t$ 的概率就可以用 $p: 100-p$ 表示。这是基于胜利优于其他一切可能的假设。投资者大都喜欢能赢得交易的策略，通过一个稳定的过程找出这种相对比率的结果。不过，只考虑胜负的比率是有缺陷的。因为策略的稳定性与其他的情况有关。在判断策略的优劣时，需要很多重要信息，包括获利者获利的金额大小，以及损失者损失的金额大小。例如，如果有一个策略，80% 的时候会输掉交易，但每次损失都不超过 0.1%，而获利的时候，则总是会有超过 1% 的利润，那么这个策略也是一个稳定而可获利的系统。当我们想要判断一些说起来很容易但其实很复杂的概念，比如“稳定性”时，必须要特别说明是否有个人偏好的存在。通常偏好都是不明确的，因此常常会导致误解。这种很个人化的偏好，在很大程度上是因人而异的。

如果 $P_t >$ 中值，$P_{t+1} < P_t$ 的概率应该会大于 1；类似地，如果 $P_t <$ 中值，则 $P_{t+1} > P_t$ 的概率也会大于 1。这里连续性的假设是非常重要的，实际上价格序列是非连续的，但我们还是可以得到接近于连续情况下的结果。如果想要了解离散分布下的一些问题，可以重新回顾第 4 章的内容。

现在出现了两个问题：

（1）在交易中可以利用这些概率吗？

- 针对的是平稳随机过程中的人造数据。
- 针对的是实际的股票数据，并使用局部数据所定义的矩（有条件地近似于平稳随机过程）。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

(2) 如何定义文中的“反转”概念?

- 向中心反转需要将上述概率予以修改，变成类似 75% →50% 与 25% →50% 的情况。
- 向偏离中心的方向反转——以便超过中心的情况被允许，并且前面用到的概率表示方式就显得比较贴切了。

在（1）和（2）中，哪一个说法是恰当的（在本章中，恰当可以解释为“对于价格序列中反转的分析与解释是有用的”）？应该如何描述反转的特征呢？是否可以用向某一方向移动的百分比（与分布无关的值）来表示？是否可以用价格移动量的期望（平均）值（这需要把假设的价格分布综合起来，这是一个与分布有关的值）作为反转现象出现的标志?

前面讨论的两个问题对于一个交易系统来说，都非常重要。在一个传统的交易策略中，按照对价格移动的方向与大小的预测进行交易，总利润就会受到反转数量的期望值的影响。模型越是合理，价格移动的期望值越大，利润的期望值也就越大。利润的波动率在一定程度上，是由价格移动的盈亏比例决定的。以相同的无条件价格分布为例，盈利的机会越大，利润的方差就越低，产生了较多的盈利交易，以及较少的亏损交易，总体上就能产生利润。止损规则的存在与交易规模的大小会对交易结果造成显著影响，相比之下，进行一般化的绝对性陈述是不明智的行为。

后面所提到的观察结果是很有价值的。假设我们只取那些可获利的交易。每日的利润变动一般来说都相当可观。通过爆米花过程模型，按照真实的数据进行交易，显示盈利的可能性为 75% 的交易策略，其夏普比率也超过了 2，但实际上盈利的机会只有 52% 。

反转可以定义为从当日价格向分布中心方向的任何移动，包括超越中值的情况。这个方案，将超过中值的价格向下的移动当成反转移动——就算移动到低于中值的范围，也就是超越中值的状况也同样包括在内。同样，一个低于中值的价格向上的移动也是反转。

7.2.1 反转移动的频率

对于任何大于中值的价格，

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

$$P_t = p_t > m:$$

$$\Pr[P_{t+1} < p_t] = F_p(p_t)$$

其中 $F_p(\cdot)$ 表示概率的分布函数，假设所有的价格都是根据这个分布而产生出来的。（这是独立性假设的一个直接结果）在这种情况下，反转发生率的衡量方式可以表示为：

$$\int_m^{\infty} F_p(p_t) f_p(p_t)\,\mathrm{d}p_t = \frac{3}{8}$$

其中 $f_p(\cdot)$ 表示价格分布的密度函数。因此，再考虑价格小于中值的情况，即 $P_t < m$，我们可以说，反转发生的可能性有75%。这一点，我们在第4章中，已经详细讨论过并加以证明。

前面我们曾提到，反转移动的比例是对价格序列特性的一种描述，并且这种描述是有用的。通过对第4章中75%比率的说明，我们知道分布形式并不会造成反转移动比例的不同。因此，对于一个利用反转现象的交易系统来说，波动率低的股票与波动率高的股票，都会出现出相同的机会。而且，倾向于出现相当大规模移动（在统计学上，称为异常值）的股票，与没有这种倾向的股票，都会呈现出相同的反转机会。这个结果的实际意义在于它说明了反转移动的比例并不是一个随机分布的函数。因此，当把某个特殊的数据加进模型中，用来计算价格时，即使假定这个分布是特殊的，情形也不会变得更复杂。而且，当分析真实的价格序列时，如果在反转移动比例中观察到了一些差异，那么，毫无疑问地，一定是时间结构上的差异造成的，而不是由于随机分布成分的不同。

7.2.2 反转数量

前面才讨论过反转发生的概率，现在如果还想要计算出某个大于中值的价格的反转数量的期望值是多少。可以使用下面这些可能的计量方式：

1. $E[P_t - P_{t+1} \mid P_t > P_{t+1}]$：反转确实发生时，反转数量的期望值。
2. $E[P_t - P_{t+1} \mid P_t > m]$：平均反转数量。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

3. $E[P_t - P_{t+1} \mid P_t > P_{t+1}]\Pr[P_{t+1} < P_t]$：反转数量的总期望值。

[注释] $P_t > m$ 是一个前提条件。案例 1 与案例 2 之间的区别是，案例 2 还包括了额外 25% 的情况，在这 25% 的情形下，虽然 $P_t > m$，但是反转并没有发生，也就是 $P_{t+1} > P_t$。案例 1 并不包含这种情形，而是只包括反转移动的情况。

如果我们利用案例 1 定义系统中“纯粹”反转的数量，那么案例 2 就可以被认为是系统中整体“揭示”的反转数量。采用这种术语，可以设想一个揭示反转数量为零或为负数的系统，而纯粹的反转数量永远都是正数（除了不令人感兴趣的和退化的例子）。

对某个小于中值的价格移动，也可以用相似的方式对其特征进行描述。

纯粹反转 纯粹反转期望值的定义为：

$$E[P_t - P_{t+1} \mid P_{t+1} < P_t, P_t > m] \times \frac{1}{2} + E[P_{t+1} - P_t \mid P_{t+1} > P_t, P_t < m] \times \frac{1}{2}$$

这两项分别对应到（a）大于中值的价格向下移动的状况，以及（b）小于中值的价格向上移动的状况。将这两种情况区分开来是很重要的，因为这两种情况的期望值是不同的；只有在对称分布时，才是一样的。现在只考虑第一种情况。在图 7-1 中，对这个比较有趣的情况，定义一个条件分布，在图中用灰色的部分表示。对于任何一个特定的价格 $p_t = p_t > m$ 来说，其反转数量的期望值等于 p_t 与条件分布期望值的差：

$$p_t - E_{P_{t+1} \mid P_{t+1} < p_t}[P_{t+1}] = p_t - \int_{-\infty}^{p_t} p_{t+1} f_{P_{t+1} \mid P_{t+1} < p_t}(p_{t+1})\,dp_{t+1}$$

当 $P_{t+1} < P_t$ 时，P_{t+1} 的条件分布密度函数就是无条件密度函数（以独立性假设为基础），这个比例因子就是 1 减去在条件之外的子集合概率。因此，反转的期望值便是：

$$p_t - \frac{1}{F_P(p_t)} \int_{-\infty}^{p_t} p f_P(p)\,dp$$

现在我们感兴趣的是 $P_t = p_t > m$ 的值所对应的反转数量期望值：

$$E_{P_t > m}[P_t - E_{P_{t+1} \mid P_{t+1} < P_t}[P_{t+1}]]$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

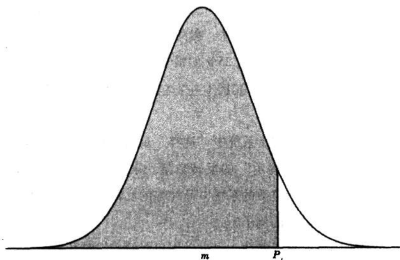


图 7-1　一般的价格分布

$$=\int_{m}^{\infty}\left(p_t-\frac{1}{F_P(p_t)}\int_{-\infty}^{P_t}pf_P(p)\,dp\right)f_{P_t|P_t>m}(p_t)\,dp_t$$

将 $P_t > m$ 引入 $f_{P_t|P_t>m}(p_t)=f_p(p_t)/(1-F_p(m))=2f_p(p_t)$，当 $P_t>m$ 时，（每天的）纯反转数量的总期望值便是：

$$E[P_t-P_{t+1}\mid P_{t+1}<P_t, P_t>m]$$

$$=2\int_{m}^{\infty}\left(p_t-\frac{1}{F_P(p_t)}\int_{-\infty}^{P_t}pf_P(p)\,dp\right)f_P(p_t)\,dp_t$$

以类似的方法，分析原来期望值公式中的第二项，推导出：

$$E[P_{t+1}-P_t\mid P_{t+1}<P_t, P_t>m]$$

$$=2\int_{\infty}^{m}\left(\frac{1}{1-F_P(p_t)}\int_{P_t}^{-\infty}pf_P(p)\,dp-P_t\right)f_P(p_t)\,dp_t$$

将这两个结果（分别乘以 1/2）相加，纯粹的反转期望值为：

$$\int_{m}^{\infty}\left(p_t-\frac{1}{F_P(p_t)}\int_{-\infty}^{P_t}pf_P(p)\,dp\right)f_P(p_t)\,dp_t$$

$$+\int_{-\infty}^{m}\left(\frac{1}{1-F_P(p_t)}\int_{P_t}^{-\infty}pf_P(p)\,dp-p_t\right)f_P(p_t)\,dp_t$$

如果能简化这个数学公式就更好了。在后面的示例 1 中，我们就看到了

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

这个公式在正态分布下的可能的简化形式。可以将这里的二元积分化简为一个一元积分：这样还是很难找到一个封闭形式的解。正态分布函数是一个对称的分布函数。如果分离点不是中值，而是平均数，能简化结果吗？当然，采用对称密度函数这样一个假设，会导致 P_t 的两个直接项互相抵消；事实上，这两个部分的和是相等的。采用 Fubini 定理，交换积分的顺序，会不会同样适用呢？我们可以知道，结果一定是正值！虽然目前无法得到一个封闭形式的解，但是针对某些特定示例的计算，都是很简单的（参见后面的示例）。

显示反转：显示反转期望值定义如下：

$$E[P_t - P_{t+1} \mid P_t > m] \times \frac{1}{2} + E[P_{t+1} - P_t \mid P_t < m] \times \frac{1}{2}$$

考虑公式中的第一项：

$$E[P_t - P_{t+1} \mid P_t > m] = E[P_t \mid P_t > m] - E[P_{t+1} P_t > m]$$

$$= E[P_t \mid P_t > m] - E[P_{t+1}] \text{（根据独立性假设）}$$

$$= E[P_t \mid P_t > m] - \mu$$

其中 μ 表示价格分布的平均值。同样的，公式中的第二项，也可以化简为 $E[P_t - P_{t+1} \mid P_t > m] = \mu - E[P_t \mid P_t > m]$。我们可以看到，公式中的两个项，都有相同的值，如下所示：

$$u = E[P_t] = \int_{-\infty}^{\infty} p f_p(p)\,\mathrm{d}p$$

$$= \int_{-\infty}^{m} p f_p(p)\,\mathrm{d}p + \int_{m}^{\infty} p f_p(p)\,\mathrm{d}p$$

$$= \frac{1}{2}E[P_t \mid P_t < m] + \frac{1}{2}E[P_t \mid P_t > m]$$

因此：

$$E[P_t \mid P_t > m] - u = E[P_t \mid P_t > m] - \frac{1}{2}E[P_t \mid P_t < m] - \frac{1}{2}E[P_t \mid P_t > m]$$

$$= \frac{1}{2}E[P_t \mid P_t > m] - \frac{1}{2}E[P_t \mid P_t < m]$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

$$= \frac{1}{2}\mathrm{E}[P_t \mid P_t > m] + \frac{1}{2}\mathrm{E}[P_t \mid P_t < m] - \mathrm{E}[P_t \mid P_t < m]$$

$$= \mu - \mathrm{E}[P_t \mid P_t < m]$$

因此，显示反转的总期望值的另一种等效的表达方式为：

$$\text{显示反转的总期望值} = 0.5 \times (\mathrm{E}[P_t \mid P_t > m] - \mathrm{E}[P_t \mid P_t < m])$$

$$= \mathrm{E}[P_t \mid P_t > m] - \mu$$

$$= \mu - \mathrm{E}[P_t \mid P_t < m]$$

除了我们不感兴趣的、衰退的情况外，结果都是正值（针对连续分布）。

［注意1］这个结果为情况1提供了一个下限，情况1未包含 $\{P_{t+1}: P_{t+1} > P_t, P_t > m\}$ 和 $\{P_{t+1}: P_{t+1} < P_t, P_t < m\}$ 的情形，这两种情形的实际反转值是负数。

［注意2］在计量反转数量的时候，希望计量的方式对于位置平移的情况具有不变的特性。如果价格增加了一美元，那么以价格为单位的反转数量应该没有改变。可以很容易看出，显示反转期望值的公式具有不因位置而改变的特点：分布沿着轴移动，平均值与中值的改变幅度是相同的。从方程式中看出纯粹反转的不变性，不是那么容易。但是，考虑图7-1描述的情况，可以看出纯粹反转的不变性。

一些特定的示例

示例1 价格为正态分布。如果 X 是正态分布，其均值为 μ，方差为 σ^2，那么 $X < \mu$ 的条件分布就是正态分布的左半边。总显示反转期望值是 0.8σ（根据约翰逊与克茨著作中的方法，计算正态分布的中值；对于一半的正态分布来说，结果就是 $\mathrm{E}[X] = 2\sigma/\sqrt{2\pi}$）。因而，价格分布发散性越大，显示反转的期望值就越大：这个结果跟我们的直觉以及期望一致。

根据标准正态分布中1 000个随机样本的结果，样本值为0.8，与理论值非常接近。图7-2显示了样本的分布。

对于正态分布，纯粹反转的一般结果还可以再简化。首先，如前面所说的，因为密度函数是对称的，所以 P_t 项互相抵消得到：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

$$-\int_{m}^{-\infty}\left(\frac{1}{F_P(p_t)}\int_{-\infty}^{P_t} pf(p)\,\mathrm{d}p\right)f(p_t)\,\mathrm{d}p_t + \int_{m}^{-\infty}\left(\frac{1}{1-F_P(p_t)}\int_{p_t}^{\infty} pf_P(p)\,\mathrm{d}p\right)f_P(p_t)\,\mathrm{d}p_t$$

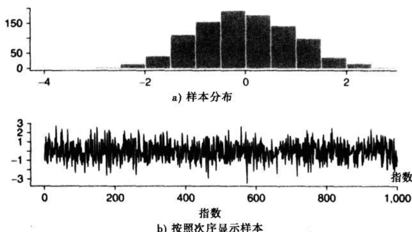


图 7-2 随机的“价格”序列

（在这里去掉 f 与 F 的下标，因为不需要区分条件分布和无条件分布：只使用无条件的价格分布）。约翰逊与克茨对缩短的正态分布的矩给出了结论。需要特别注意的是，缩短的正态分布的期望值可以表达为：

$$\int_{-\infty}^{P_t} pf(p)\,\mathrm{d}p = -\frac{f(p_t)}{F(p_t)}$$

和

$$\int_{P_t}^{\infty} pf(p)\,\mathrm{d}p = \frac{f(p_t)}{1-F(p_t)}$$

因此，纯粹反转的期望值就变成了：

$$-\int_{m}^{-\infty}\frac{1}{F(p)}\frac{-f(p)}{F(p)}f(p)\,\mathrm{d}p + \int_{-\infty}^{m}\frac{1}{1-F(p)}\frac{f(p)}{1-F(p)}f(p)\,\mathrm{d}p$$

$$=\int_{m}^{\infty}\left(\frac{f(p)}{F(p)}\right)^2\mathrm{d}p + \int_{-\infty}^{m}\left(\frac{f(p)}{1-F(p)}\right)^2\mathrm{d}p$$

检查一个对称的密度函数，可以发现，公式中相加的两个表达式是相等的。采用代数的方法，运用 $f(m-\varepsilon) = f(m+\varepsilon)$ 和 $F(m-\varepsilon) = 1 - F(m+\varepsilon)$ 这两个式子，然后进行变量之间的简单转换，$q = -p$，立即得到结果。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

$(1-F(x)[87])/f(x)$就是"米尔斯比率"（Mills'ratio）。因此，正态独立同分布（iid）模型的纯粹反转期望值，就是米尔斯比率的平方取倒数的积分乘以2，其中取实线一半的积分：

$$2\int_{-\infty}^{m} M(p)^{-2}\mathrm{d}p$$

图7-2b显示了本节一开始涉及的时间序列随机样本。对于这个序列：从中值上面往下移动的数量，$\{x_t : x_t > 0 \cap x_{t+1} < x_t\}$，是351；从中值下面往上移动的数量，$\{x_t : x_t < 0 \cap x_{t+1} > x_t\}$，为388；反转移动的比例是$100\% \times (351+388)/999 = 74\%$（因为两个数据（$x_t$，$x_{t+1}$）只能算成一组数据，所以分母减去一。当然，没有值可以跟x_{1000}匹配）。

时间序列的显示反转：图7-3显示了（a）从一个大于中值的数值，移动到下一个数值的移动量，$\{x_t - x_{t+1} : x_t > 0\}$，以及（b）从一个小于中值的数值，移动到下一个数值的移动量，$\{x_{t+1} - x_t : x_t < 0\}$，这两种状况对应的每日"价格"移动量的分布图。在意料之中的是，对于这样一个数量比较多的样本来说，这两种情况的分布图，看起来非常的相似（事实上，这两个显然很相似的图形，不管是取总和，还是计算样本的期望值，结果都相当的一致）；这些移动量的总和平均是$794/999=0.795$，这个值非常接近理论上的期望值，0.8。实际上虽然有一点不一致，但还是在预期之内。

这个时间序列的纯粹反转：图7-4a是从大于中值的数值，向下的移动量，$\{x_t - x_{t+1} : x_t > 0 \cap x_t > x_{t+1}\}$，图7-4b是从小于中值的数值，向上的移动量，$\{x_{t+1} - x_t : x_t < 0 \cap x_t < x_{t+1}\}$。图7-4显示了（a）和（b）对应的每日"价格"移动量分布。这些柱状图是以图7-3为基础，去掉小于零的数据之后，所得到的简化版本。纯粹反转的总值是$957/999=0.95$。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

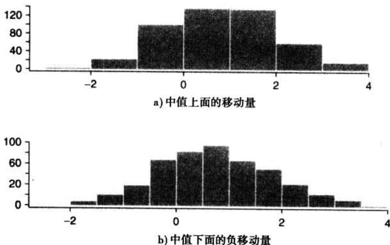


图 7-3 随机“价格”序列：每日价格反转移动的分布

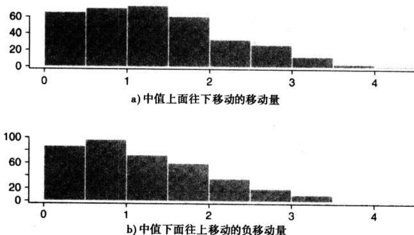


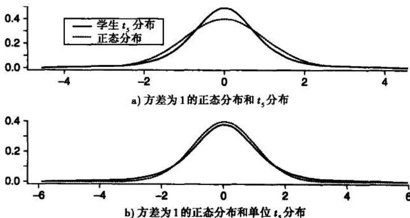
图 7-4 随机“价格”序列：每日价格纯粹反转移动的分布

在示例 5 中，我们还会更进一步分析示例中的数据。

示例 2 价格属于是自由度为 5 的 t 分布。蒙特卡罗实验得出显示反转期望值为 0.95（对应单位参数为 1 的 t_5 分布）。注意，这个值大于单位参数为 1 的（标准）正态分布对应的值（0.8）。值比较大的原因是由于学生 t 密度函数同正态分布相比，其中心收紧，而尾巴拉长：比较厚的尾巴意味着，在远

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

离中心的地方也会发生大概率事件。回想 t 分布的方差是等比例乘以 $dof/(dof-2)$，其中 dof 表示自由度，而参数为 1 的 t_5 分布方差为 $\frac{5}{3}$。换句话说，参数为 $\frac{3}{5}$ 的学生 t_5 分布的方差为 1，显示反转就是 $0.95\times\sqrt{3/5}=0.74$，这个值小于标准正态分布。



a) 方差为 1 的正态分布和 t_5 分布

b) 方差为 1 的正态分布和单位 t_5 分布

图 7-5　正态分布与 t 分布的比较

从图 7-5 中，可以看到更详细的比较。

纯粹反转：参考示例 1 中的评论。

示例 3　价格服从柯西分布。由于柯西分布的矩没有定义，反转期望值的计量方式也没有定义，所以在现阶段，它是一个不能得出很多结论的示例。如果加上一些限制条件，例如对柯西分布进行删节，在 t 分布的研究基础下，这将会是一个有趣而且活跃思维的例子，因为参数为 1 的柯西分布就是自由度为 1 的 t 分布。

示例 4　一个依赖经验的实验。股票 AA 在 1987 ~ 1990 年期间，每日收盘价格（除息后的数据）显示在图 7-6 之中。显然，每天的价格不是完全独立的，也不能假设服从相同的分布。通过价格位置与价差的变化，对价格序列

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

进行局部调整，可以减少这些违反假设条件的样本。即使这样，我们不能希望会出现75%的反转结果：很难用一个方式来描述数据序列的相关性结构。根据前面所涉及的计量方式，更感兴趣的是，在现实中会出现多少反转现象（实际结果与理论结果不相同，可以归咎于违反了某些假设条件）。

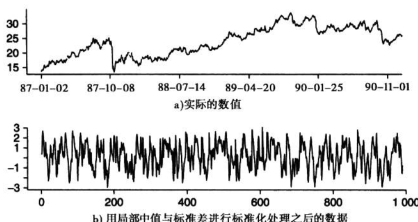


图7-6 股票AA的每日收盘价格（除息后的数据）

用每天的价格先减去一个局部估计值（可以采用平均值或者是中值），然后再除以一个局部标准差的估计值，将价格序列转换成一个标准化的序列。这两个局部估计值，都是用近年来的历史数据，按照指数加权移动平均的方法计算得来。用这种方式模仿实时的操作流程（见第3章）。图7-6显示了标准化之后，交易周期为10个交易日的价格序列，估计值是中值。与后面的图7-8比较，图7-8只对价格序列的位置进行了调整。

对于一个调整了位置，没有进行方差标准化的序列来说，反转移动的比例是58%，比理论上75%的结果低很多。请注意，零作为这个调整过序列的中值。从结构上来说，中值应该要接近于零；更重要的是，这种选择使程序操作起来更方便。我们做更多的实验，采用不同的加权方案，使用20个交易日作为有效的周期，然后用局部平均值取代局部中值来作为位置的估计值，得到的结果也都类似。反转移动的比例都在55%～65%的范围之间。这个结果说明，只对一个序列的位置进行局部的调整，就会有一个或多个假

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

设条件无法满足。

图 7-7 显示了价格变化的分布，在预测价格将会往下回归的那段期间，价格变化等于收盘价格减下一个收盘价格；在预测价格将会往上回归的那段期间，价格变化等于下一个收盘价格减收盘价格。价格变化是根据原始的价格序列（而非调整中值后的结果）计算出来的。这是价格才能真正决定一个交易策略的结果。因此，图 7-7 显示原始价格的损益分布图，根据股价相对于局部平均价格的移动情况来决定交易策略。图 7-7a 显示预测价格将从局部中值之上向下反转，实际交易的损益分布情形，图 7-7b 则显示预测价格将从局部中值之下向上反转，实际交易的损益分布情形。可以很清楚地看到，一般任何一个趋势都不会获利。总的来说，962 笔价格在 15 ~30 美元之间的股票的交易利润是 -31.95 美元（其中有 28 天局部中值调整为零。由于要计算局部中值与标准差，$2\times k=20$ 天也没有被列入考虑）。

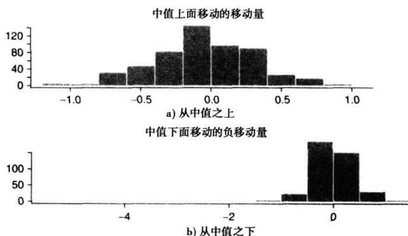


图 7-7 股票 AA 围绕局部中值每日的价格移动

图 7-8 显示了价格减去局部中值之后的分布。采用第 7.2 节介绍的方法来计算的话（中值为零），总显示反转为每天 597.79/990 = 0.60 美元。在上一段中，采用原始价格计算出来的实际利润为 -31.95 美元，说明一些无法满足假设条件的情况（由于结构无法模型化），影响了这个例子中的显示反转期望值。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

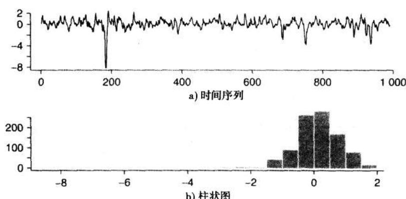


图 7-8 用局部中值调整过的股票 AA 收盘价格

对调整过中值的序列进行分析，可以清楚地看到“缺少的结构”产生的影响。这个序列总显示反转数量（根据调整过的序列给出的交易信号，计算反转数量）是 70.17 美元。从局部位置调整的数据序列中，得到的显示反转值不到独立同分布模型的八分之一（在实际交易系统的研究中，剔除离群点是一种必要的练习。在实际操作中，风险控制程序会屏蔽巨大的离群点）。示例 5 中尝试着想要解答这一问题，那就是，即使调整了局部数据，还是会损失多少反转机会。

图 7-9 显示了样本自相关性与偏自相关性的估计值：显然，在局部调整序列中，1 天与 2 天的序列结构具有非常强烈的相关性。毫无疑问的，在独立性的假设前提下，这也是导致部分的（大部分的?）实际显示反转值与理论期望值之间存在很大偏差的原因。

纯粹反转：如果用中值调整过的价格序列，在 562 天内得到的总纯粹反转是 191.44 美元（对于实际的时间序列，我们还没有一个封闭形式的解，应用到价格移动序列中。不过，我们知道它一定会超过显示反转理论估计值，597.79 美元）。按照原始的价格序列，纯粹反转（运用调整过的序列的交易信号）只有 60%，就是 117.74 美元。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

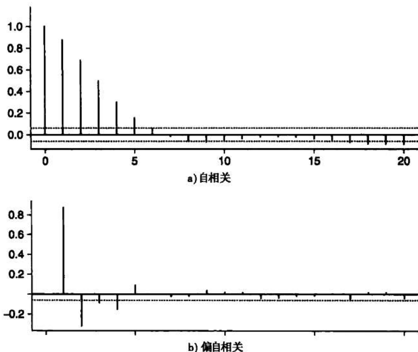


图 7-9 股票 AA：局部中值调整过的收盘价格

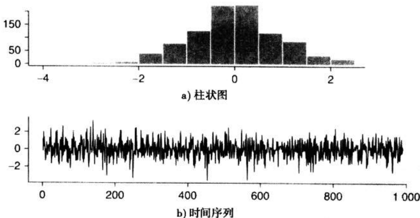


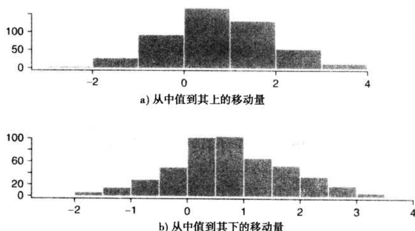
图 7-10 运用局部中值调整图 7-2 的随机价格序列

示例 5 这是对示例 1 中数据的一个后续的分析。在示例 4 中，用局部中值进行调整的操作方法，把 75% 的结果应用到真实但非平稳的数据序列之中。如果能够知道这种调整的操作程序对于一个实际上处于平稳状态的序列而

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

言，所代表的含义也是很有趣的事情。当违反其他假设条件时，例如序列相关性，这样的知识可以帮助我们决定实际的调整所产生的影响。

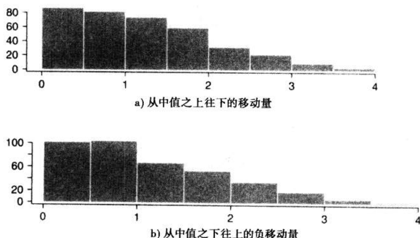
从图 7-10 到图 7-12，是将图 7-2 到图 7-4 的数据，进行局部中值调整之后得到的数据。（局部中值是用前 10 个数据计算出来的。）汇总统计值和在括号中显示的示例 1 中原来的分析值，分别是：77%（74%）的反转移动；总纯粹反转 =900（952）；总显示反转 =753（794）。



a) 从中值到其上的移动量

b) 从中值到其下的移动量

图 7-11 随机价格序列，用局部中值方式描述与图 7-3 相似的图形



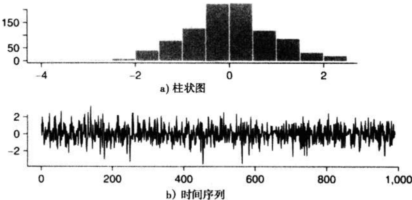
a) 从中值之上往下的移动量

b) 从中值之下往上的负移动量

图 7-12 随机价格序列，用局部中值方式描述与图 7-4 相似的图形

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

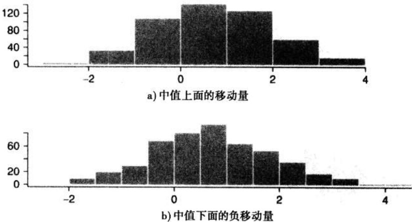
从图 7-13 到图 7-15 所描述的数据分析是很中肯的。利用中值调整过的序列，判断一个点是位于局部中值的上面还是下面。在计算反转数量的时候，则采用了与前面段落不同的方式，直接运用原始而未经调整过的数据序列。这种做法在实践中比较有意义。任何选定的模型都有可能产生交易的信号，但实际的市场价格才真正决定着交易的结果。有趣的是，纯粹反转增加到 906，但是这个值仍然低于未调整过的序列结果 952。显示反转降至 733。



a) 柱状图

b) 时间序列

图 7-13　随机价格序列，交易信号产生于局部中值调整过的序列，而反转数量来自于原始的序列



a) 中值上面的移动量

b) 中值下面的负移动量

图 7-14　随机价格序列，交易信号产生于局部中值调整过的序列，而反转数量来自于原始的序列

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

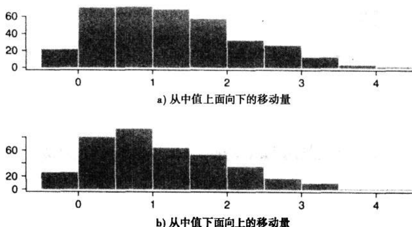


图 7-15 随机价格序列，交易信号产生于局部中值调整过的序列，而反转数量来自于原始的序列

在图 7-15 中，还有一些很有趣的事情。请注意，“从中值之上向下移动的量”出现了负数，从逻辑上来说，这是不可能的！交易的信号是根据局部中值调整过的价格序列计算出来，但是价格的移动量是根据原始而未经过调整的序列计算出来。负移动量相对比较少，是有可能出现。在这个特别的样本中，即尽管纯粹反转出现了负数，但总纯粹反转实际上却增加了，这看起来很奇怪。原始序列的正移动量，在大小上比调整过的序列还要大，便出现了这种现象。

示例 6 价格服从对数分布。如果 $\log X$ 是正态分布，其平均值为 μ，方差为 σ^2，那么 X 就是对数分布，平均值为 $\mu_X = \exp(\mu + 1/2\,\sigma^2)$，方差为 $\sigma^2 = \exp(2\mu + \sigma^2)\exp[(\sigma^2) - 1]$，中值为 $\exp(\mu)$。应用约翰逊与克茨的著作第 1 册 129 页的结论，对于一个删节的对数分布来说，总显示反转期望值是：

$$\exp(\mu + \sigma^2/2)[1 - 2\Phi(-\sigma)]$$

$\Phi(\cdot)$ 表示标准正态累积分布函数。在本章最后的附录 7A 之中，论述其他一些细节。图 7-16 显示了 $\mu = 0$ 与 $\sigma = 1$ 的一个对数分布的 1 000 个随机样本，对应的柱状图以及时间序列（中值减去 1，使得分布的中心为 0）。样本的显示反转期望值是 1.08（理论值是 1.126）。按照示例 5 中处理时间序列的方式，处理样本数据，则样本的纯粹反转估计值变成 1.346。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

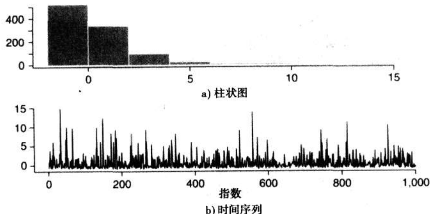


图 7-16 对数分布得到的随机样本（减去中值）

7.2.3 用分位数而不是中值，作为判断移动的标识

到目前为止，对所有移动的分析，都是集中在判断价格是位于中值之上还是之下。从图 7-17 显示，对于正态分布与学生 t 分布，中值是一个合理的点。例如，当价格超过 60%（在对称的情况下，还包含未超过 40% 的情况）时，如果我们考虑价格移动的子集合，从今天到明天价格变化的期望值，与价格超过中值时所得到的期望值相比，要小一些。

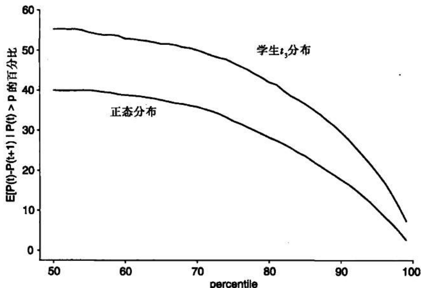


图 7-17 纯粹反转期望值，与价格条件分布的百分比

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

如果检验的是具有相关性的序列，可以预测这个结果不一定正确。交易策略还要考虑交易成本，这里的分析也忽视了这点。

7.3 非平稳过程：不均匀的方差

在7.2节中，检验中采用的数据都是严格的“平稳”、“独立”、“同分布”过程，在本节中，会放宽这些非常严格的假设条件。在这里，将会把纯粹反转与显示反转的计量方式，推广到“独立”、“同分布”，但“非平稳”（方差不均匀）的过程中。关于“75%规则”的推广见第4章。

假设价格是根据一个分布家族中的一个特定分布，独立产生出来的。这个分布家族是固定的，但每天的价格所依据的分布是不确定的。分布家族中不同的分布只是方差不同。方差序列的特性就决定了价格序列。

7.3.1 顺序结构的方差

根据某个已知但不同的方差，每天产生出一个正态价格序列，那么7.2节的结论，就可以直接应用到正态序列之中。回顾一下，将一个正态价格序列与理论上的期望值进行比较（这与我们在4.3节中的做法一样，但4.3节中不需要正态分布）。当然，可以计算出实际的纯粹反转与显示反转值。然而，我们不能得出任何结论，建立一个能够应用反转期望值的交易规则。

存在一种做法，就是观察每天方差的范围，以及在每天的值中呈现出的任何系统性的模式。在极端的情况下，如果能预测明天的方差，那么只要对75%规则稍作修改，就可用于计算纯粹反转与显示反转的值。如果相同（或者在实践中非常相似）的环境（今天的方差与明天的方差）经常发生，那么根据概率分布计算的期望值，会相当有用。只要方差簇（variance cluster）能描述金融数据序列，这种做法就有现实意义。关于这模型化的讨论见第3章。

用一个特别简单的例子作为预测方差的示范。假设在价格序列中，非平稳的方差只在星期五出现。星期五这一天的方差是一周之内其他日子的两

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

倍。在这个例子中，不需要修改规则或计算反转期望值：将 75% 规则、计算纯粹反转期望值和显示反转期望值，应用到价格序列中时，忽略星期五就行。回想一下，我们假设每天价格是独立的，因此从价格序列历史数据中有选择性的删除一部分数据，并不会对结果的有效性造成影响。在实践中，序列的相关性会影响这个简单的解决办法。而且，如果再努力一些，得到一个更通用的解法，为什么还要舍弃星期五的交易数据呢?

7.3.2 顺序无结构的方差

在第 4 章第 3 节中详细分析过这种情况。对于正态 – 逆伽马示例（见图 4-2），反转期望值的计算结果：实际的显示反转是每天 206.81/999 = 0.21；纯粹反转是每天 260.00/999 = 0.26。请注意，纯粹反转与显示反转的比率，0.26/0.21 = 1.24，大于 7.3 节中正态分布（示例 1）的结果，0.94/0.8 = 1.18。

7.4 序列的相关性

在 7.3 节的示例 4 中，对股票 AA 的分析显示了，价格序列，实际上是局部中值调整过的序列，呈现强烈的一阶自相关性，以及较弱但仍旧值得关注的二阶自相关性。我们认为独立同分布序列反转期望值的理论结果（75%），与（应用中值调整过的）序列实际计算结果（58%）不相同，可能与序列的相关性有很大的联系。与序列相关性的扩展理论见第 4 章第 4 节。我们用第 4 章的示例作为本章的结束语。

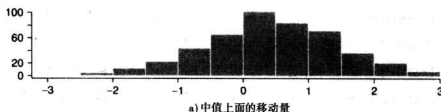
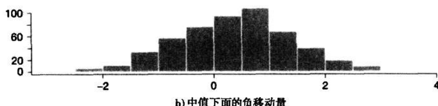


图 7-18 应用局部中值调整之后，自相关模型的分析

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net



b) 中值下面的负移动量

图 7-18　（续）

示例 7　根据一阶自回归模型，其中序列的相关系数 $r=0.71$，而随机数据服从正态的分布（参见第 4 章的示例 1），产生 1000 个数据。图 4-4 是时间曲线图与样本的边际分布。这个序列的反转移动比例是 62%；总显示反转是 315，而总纯粹反转是 601。忽略序列的相关性，并使用样本边际分布，计算 7.2.2 节中的结果，得出理论上的显示反转是 592，这个值是实际值的两倍。图 7-18 显示了这个分析的图形。

应用局部中值调整之后的序列（周期为 10 天）呈现在图 4-4 中；反转估计值的图形呈现在图 7-19 中。这个调整后的序列的反转移动比例更大一些，为 65%（这是与原始序列 62% 相比）。总显示反转是 342（未调整的序列总显示反转为 315）；总纯粹反转是 605（未调整的序列总纯粹反转的值是 601）。

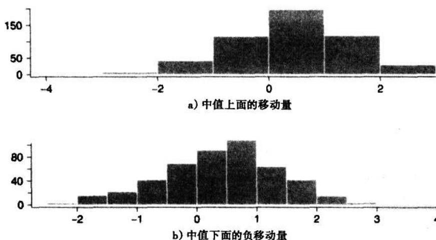


图　7-19

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

附录 7A 在示例 6 中对数分布的一些细节

$$Y = \log X \sim N[\mu_Y, \sigma_Y^2]$$

规定：

$$E[X] = \mu_X = \exp(\mu_Y + \sigma_Y^2/2)$$

$$V[X] = \sigma_X^2 = \exp(2\mu_Y + \sigma_Y^2)[\exp(\sigma_Y^2 - 1)]$$

$$中值 = \exp(\mu_Y)$$

定义 $Z = X$，在 X_0 进行删节（也就是，Y 在 $Y_0 = \log X_0$ 的地方进行删节），那么（根据约翰逊与克莱的著作第 129 页）：

$$E[Z] = \mu_Z = \mu_X \frac{1 - \Phi(U_0 - \sigma_Y)}{1 - \Phi(U_0)}$$

其中：

$$U_0 = \frac{\log X_0 - \mu_Y}{\sigma_Y}$$

总显示反转期望值可以写成 $E[X|X > m] - E[X]$。现在，$E[X|X > m] = E[Z]$，其中 $X_0 = m = \exp(\mu_Y)$。这样，U_0 就简化为 0，而：

$$\mu_z = \mu_X \frac{1 - \Phi(-\sigma_Y)}{1 - \Phi(0)} = 2\mu_X[1 - \Phi(-\sigma_Y)]$$

因此，总显示反转期望值就是：

$$2\mu_X[1 - \Phi(-\sigma_Y)] - \mu_X = \mu_X[1 - \Phi(-\sigma_Y)]$$

$$= \exp(\mu_Y + \sigma_Y^2/2)[1 - \Phi(-\sigma_Y)]$$

特殊情况：$\mu_Y = 0, \sigma_Y = 1$。那么，$\mu_X = \sqrt{e}$，$\sigma_X^2 = e(e-1)$，X_0 = 中值 = $e^0 = 1$。现在，$U_0 = \log X_0 = 0$，因此：

$$\mu_z = \sqrt{e}\frac{1 - \Phi(-1)}{1 - \Phi(0)} = 2\sqrt{e}[1 - \Phi(-1)]$$

根据标准统计表中（参见在约翰逊，克莱与巴拉克瑞兰著作中的参考书目），$\Phi(-1) = 0.158\ 65$，因此删节的对数分布的平均值 μ_z（其中 $\mu_Y = 0$，$\sigma = 1$）是 2.774。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

第 8 章 诺贝尔的困惑

机遇总是垂青那些有准备的人。

——法国科学家 路易斯·巴斯德

8.1 导论

在本章中，我们将检查一些基于统计套利但产生负面结果的案例。在运行一项投资策略时，尽管进行了风险过滤，也有止损规则，但意外还是会经常发生。在第一个示例中，检查一个单一的匹配交易，它一直都表现出了书中所讲的反转现象，直到后来有了根本的变化，即一个收购的公告，形成了一个断点。接下来，我们将讨论 1988 年的信用危机对国际经济的发展产生的双重影响：一方面，它为股票市场引入一个新的风险因素，即影响股票临时价格的“公司债务信用评级”。另一方面，它将一个获利的年度（到 5 月）变成了赔钱的年度（到 8 月）。接下来，我们考虑对冲基金、共同基金、退休金基金等这类基金的赎回量要达到多大，才会造成股价动态变化，以致产生临时性的崩溃，从而影响统计套利的绩效。接下来，我们讨论《公开披露条例》。最后，在所有关于统计套利的不利影响的讨论中，我们将再次讨论第 5 章的主题，以便消除一些误解，特别是要消除对负回报期间管理人员绩效相关性的误解。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

8.2 事件风险

图 8-1 显示了从 1996 年 1 月 2 日～1998 年 3 月 31 日，美国联邦家庭抵押贷款公司（FRE）与 Sunamcfica 有限公司（SAI）股票的历史价格（即除息后的每日收盘价格）。这两个价格轨迹非常接近，并具有一个非常强的向上趋势，两者之间的价差不断地扩大和缩小。

随后的分析与示例，主要集中在匹配交易上，但是结构性转移的优点，与统计套利的模型有更强的相关性。采用因素模型来预测（在股票组合中）个别股票的移动，同描述运动一样，机制比较复杂，很难解释其细节。我们尽量让分析简单，并再次提醒读者，这里所涉及的基本观点被广泛运用到了统计套利模型中。

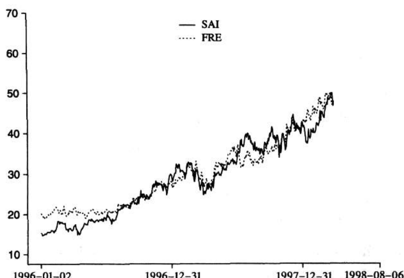


图 8-1 调整过的 FRE 与 SAI 的股票历史价格

这两个股票的每日回报没有呈现出很强的相关性，相关系数只有 0.4。然而，观察事件之间的回报，其相关性很高，相关系数上升为 0.7。在给定的风险范围内，为保证交易顺利进行，事件相关性暗示了可以预测什么内容（见第 2 章）。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

从视觉和统计学来看，可以运用一个简单的价差反转模型来进行［FRE，SAI］这个匹配交易，并且可以交易得很好。一个基本的爆米花过程模拟模型（见第 2 章）也证明了，确实可以这么做。

图 8-2 显示了从 1998 年的第二季度到 8 月 6 日，调整过的 FRE 与 SAI 价格序列。感兴趣吗？价差变动相当大，超过了历史最大值的两倍。正如之前已经提及的，价差的大小不会造成交易的损失，但价差逐渐变大的过程会带来实际的损失。价差开始在一段期间内持续增大时，当持续的时间比预期还要长，或者当局部的平均价差值与交易进场时很不相同，并产生亏损时，交易就会倾向于结束（在这个分析中，假设预测模型能够自动调整，并且没有额外的干预。如果采取监控措施，或者是有止损的规则，都有可能减少损失，但是在不改变基本面的情况下，将描述变得更复杂）。

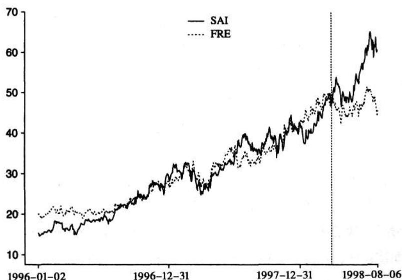


图 8-2 到 1998 年 8 月，调整过的 FRE 与 SAI 的股票历史价格

假设存在一个爆米花过程，限制其趋势（见第 3 章），使用指数加权移动平均数模型，在 2 月下旬到 4 月下旬，以及 6 月上旬到 7 月上旬之间，各自建立起一笔［FRE，SAI］之间的匹配交易。这两个场合都会带来一定的

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

损失（在7月下旬，会突然好转，成为盈利的交易）。

不断变小的价差一定能获利吗

很遗憾，实际上并没有一定会获利的方案。然而，同波动率递增相比较，波动率递减的情形，具有因不对称所带来的好处。因为在波动率递减的情形下，当局部平均值变化时，模型迟延的观点会更具优势（回想一下，当局部平均值保持不变时，波动率的变化并不是个问题。如果对波动率的预测不够迅速的话，就可能会错过获利的机会，从而产生机会成本）。

当价差的局部平均值持续向某个方向移动，并且超过了模型的调整方式，这个策略就会产生损失，因为这个交易的退出时点（零期望回报）相对于这个交易的进场时点来说，不是很好。进场时点与退出时点同时发生呈现了正确的关联。由于随着时间的变化，价差表现出了不同于交易进场时点所预测的情况，问题就产生了。如果局部波动率的预测值，比实际的波动率还大，那么目前的交易进场时点太过保守（实际交易会比较少，而且每个交易都会有比较高的期望回报）。当趋势继续向不利于模型的方向发展，这种保守主义会减少交易的次数，因而也就减少了预测不准确所带来的损失。当波动率递增，而模型的预测值低于实际值，就会产生相反的结果。

当价差的关系发生变化时，止损是一种明智的做法。可是，如果实现止损，还需要另外一个预测，即对于变化的预测。通常情况下，我们所能做到的最优情况，就是在变化发生之后，迅速将它识别出来，继而很快地描述其特征。即使这样，这份工作也是极具挑战性的（请参考本书第3章）。

根据1998年8月的情形，查看FRE-SAI交易，我们要问需要维持这个持续亏损的头寸吗？显然不需要。经理人自行决定要解开头寸的时点。但这个匹配交易什么时候才能重新回到候选的交易名单呢？FRE与SAI的相关性似乎被打破了。如果股票回报序列的一般底层结构中也被打破，那么股票便无法满足匹配交易的选择标准，因而交易行为也会停下来。如果

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

能够维持一般结构，并且价差的震荡到了一个新的水平，或者是回到了最近的历史水平，那么干扰一旦结束，就会继续选择股票进行交易，并获得收益。

1998 年 8 月 20 日（星期四），AIG 公告接管了 SAI 的消息。SAI 在星期三的收盘价为 $64\frac{3}{8}$ 美元。随后，SAI 的股票则以 25% 的溢价进行收购（市场开盘之前）。在接管之前，SAI 的价格一直往上跑，人们非常想要知道购买压力的来源。

8.3 一个新的风险因素的出现

在 1998 年夏天全球信用危机期间，对于统计套利来说，是一段有趣的时间。统计套利的绩效问题从 6 月就开始出现了。对大多数人来说，问题在 7 月与 8 月间急剧恶化。在这段时间中，人们第一次感觉到，投资者对公司的短期评价竟然会直接受到公司信用的影响。当投资者对股票持有负面看法，市场就会压低股价，而价格下降的大小，与公司的信用评级密切相关。信用评级比较低的公司的股价，下跌的幅度会比信用评级较高的公司大得多，甚至可以达到三倍的程度。这样一种戏剧化的识别方式，在此之前没有出现过，当然，在统计套利的历史中，也从未出现过。

当时，针对这种情况，出现了很多的假设，现在只有少数还站得住脚。它们想说明，在 1998 年信用评级与股票市场之间的联系，以及为什么会出现如此戏剧性的价格波动。毫无疑问，对冲基金美国长期资本管理公司（LTCM）的崩溃，以及美联储敦促那些原本不怎么热情的投资银行的救市措施，这些种种因素加剧了人们对“美国金融体系出现系统性失灵”这一流言的恐慌。这些措施是为了救市，但美联储所采取的行动加重了市场的恐慌。一个重要原因就是，各种信息、投机活动、流言，通过 24 小时播放的电视新闻频道和网络四处传播。大量的股票操盘手，以及活跃的散户，在 20 世纪 90 年代被“牛市”行情所吸引，纷纷进入了市场。由

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

于技术进步，交易也变得更加容易。这使大众通过各种渠道获取信息，也为恐惧与恐慌的滋生提供了沃土。可供决策的方法少得可怜，很多分析大都是一些简单技术分析，比随机给出结论好不了多少，这些方法充斥大众的头脑。这些情况有利于推高了市场的波动性，以及研究“高等生物会出现像旅鼠一样的行为”这一问题，但是不利于降低人们血压和胃溃疡发病几率。

当美国的股票交易市场开始受到俄罗斯债务危机的影响时，市场上便认为公司将会因信用问题而越来越多地被挤入市场，这会导致股价的下跌。随着危机的持续，股价下跌时，信用评级低的公司比起信用评级高的公司价格下跌得更快。由于这种效应不断的累积，更证实了当时的恐惧。这都使得信用市场紧缩（没有直接的论据可以证明这个与俄罗斯债务危机之间存在联系），最后导致需要花费更高的成本来维系金融市场。还有比“信用评级较低的公司应该花更大的代价”这样的看法，更合乎逻辑的看法吗？这种识别股票价值的方法，看起来很正确。很多“分析”都有相同的看法，最后导致预言成真。美国提高利率，也是俄罗斯债务危机造成的原因吗？

1998 年夏，在美国股票市场，公司债务评级变成了一个识别股价的因素。在选择投资组合时，如果忽略了这个因素，就有可能遭受损失，因为信用评级低的股票的价格，跟评级高的股票相比，其下跌的幅度是极不成比例的。在构建一组相符条件的匹配交易时，无论是根据一个普通的匹配交易策略，还是依据其他紧密的因素建立的回报预测模型，已经没有多大差别了。亏损是不可避免的。

随着一些辨识股价的模式被陆续开发出来，不同的结果对统计套利策略进行了区分，但是管理者在遭受了持久损失之后，是否继续依赖模型则使情况变得更加复杂：模型的绩效和管理者的能力（在某种程度上）混合在一起，不可能加以区分。如果不考虑管理者自身带来的影响，模拟结果指出，

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

因子模型会比纯价差模型具有更大的弹性⊖，也更容易出现正回报。

当识别出一个新的风险因素，应该做些什么呢？但采用历史回报将因素分解再重新计算，因子模型就会考虑“债务评级”，不需要采用其他行动。如果不采取任何行动会出现一些问题：比如在进化过程中，当识别出来某个因素（或者至少知道某个因素）应该怎么做呢？难道，只是等待一个新的数据周期的出现，以便评估股票暴露在这个分析因素的程度（同时，也只能接受市场对绩效评价）吗？对后者的答案很明显，绝不是简单的“不”字。一个比较合乎情理的答案需要进一步研究。一般最初的想法是：清除投资组合暴露在那个因素下的头寸。但问题是，据此准确行动在实践中存在困难。什么叫“暴露在风险中的头寸”呢？由于情绪的强烈影响，以及对盈利的渴望等需求下，“运气”扮演很重要的角色。

无因子模型又表现得怎么样？首先，确定在策略中因素风险的管理方式，然后，对债务评级的因素也运用这种方法。对于以匹配交易为基础的交易策略来说，使所有关于债务评级的匹配组合均匀化，是一种明显的补救方



注 释

⊖ 比较大的弹性源自哪里呢？通过对比基本的匹配交易策略和因子模型策略，可以得到部分的答案（通过观察评论的依据，就可以得出模拟结果的来源）。匹配交易投资组合，是由许多股票匹配交易组成的，这些股票必须符合标准的计量方式，包括产业类别、股本、市盈率，等等。假设价差（使用价格对数比率序列）符合一阶的DLM预测模型，并使用一个类似GARCH的反差定理。当价差偏离预测值的比率（按年计算）15%或更大时，就进行交易。持有股票直到模型发出一个退出信号为止。不调整交易的头寸，也不应用止损规则。因子模型采用第3章所描述的方法，为了与匹配交易模型进行比较，将最优化目标定为每年15%。交易头寸根据每天的预测值，以及风险暴露的程度，进行调整。

有证据说明信用评级与因素模型的结构因素组合，是相关的。在此基础之上，能将新的风险因素加入到模型之中，并使得模型的绩效表现不错。如果应用因子分析，从大量的股票中选出候选交易股票，其结果会受影响。对于匹配交易策略来说，也同样如此。虽然这样，按照因子模型所挑出来的股票绩效，要优于匹配交易模型挑选出的股票。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

法。也就是说，只有两只股票具有大致相同的债务评级时，才可以接受这样的股票匹配组合。信用评级高的股票与信用评级高的股票配成一对，信用评级低的股票与信用评级低的股票配成一对。要考虑到债务因素，从而可以避免在交易的过程中，股票价格出现不一致的价格波动。应该研究债务评级的很多方面的问题，具体包括交易头寸是否随信用评级递减，信用评级低的股票是否只能对其做空，或者对信用评级极低的公司是否投否决票等。

随着研究工作的推进，出现了一个重要的问题：考虑新模型的限制条件之后，对一个交易策略的表现会产生什么影响？找出交易策略模型中最可能影响绩效表现的因素，并应用到预防变化的措施之中，这将是一件值得庆贺的事情。但是，同时也应知道这种行为会对未来的绩效表现产生影响（不仅仅当因素造成了负面影响时，作为一种保护性措施）。在第 9 章中，将在更大的范围讨论“绩效崩溃”这一主题。

8.4 赎回压力

以广泛的股票为基础，并且持有头寸的基金，都有经过了完美的“设计”的赎回模式，抵御在市场突然转向，由多头转空时产生的冲击力。出售持有头寸的基金，便会对股票的价格产生不对称的压力，是一种向下的压力。假如出售的基金是由很多股票构成，而且在一定程度上持续减少，那么对价差头寸会造成负面的影响。

假设基金“以大量的股票为基础”，这意味着在反转策略的交易中，有很多的股票会受到影响——假设是一半的股票。假设多头投资组合策略与反转策略无关，也就是假设出售行为，对于反转策略的多头与空头来说，具有相同的影响力，这样会产生近似于如上所述的结果。那么，在基金赎回的压力之下，多头资产的价值会降低，而所有被影响的空头头寸，负债也会降低。一般而言，对于反转策略而言是没有影响的。

乍一看，这种说法是对的。但是，在反转策略中，对一般的价差头寸会产生什么影响呢？多头与空头的价格都会受到价格向下的影响，并且假设成

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

比例地下降，那么价差基本上就不会变化（在实践中，相对价高的股票，通常会作为卖出首选目标，以获得最大的收益。这对于一个价差交易来说，会造成片面的、负面的影响。对于价格低的股票来说，在出售时，会导致价格变化很大，使得一个以反转为基础的交易策略亏损，而不是没有盈利。在这里的讨论中，采用影响为零这样一个乐观的假设）。在价差组合中，只有多头或空头交易面临基金赎回压力时，价差才会发生变化。价差变窄，便会盈利；价差变宽，就会造成亏损。很多价差组合净收益为零。但是，那些变窄的价差会导致交易退出，赢得利润。变宽的价差则会继续变宽，因而造成更大的损失。而且，价格持续降低，导致价差模型建立新的头寸。而在连续的出售过程中，价差继续变宽，新头寸也会亏损。如果出售大量的头寸，导致出卖的行为持续的时间足够长，那么价差策略的自然交易周期便会完成，这些交易和新的交易最后亏损退出。

情况还会进一步恶化。当售出行为结束之后，有些股票会出现相似的趋势，好像存在持续性的买入压力。谁知道为什么这样，可能是反转现象的反弹！对于那些股票，会建立新的价差头寸。请记住，之前建立的一些头寸早已退出。价差模型已经等待新的交易信号很久了。对了！股价要开始上涨了。只不过卖方压力变成了买方压力，价差交易很有可能最终还是亏损（只是方向相反）。

高频率的反转策略，会在相对小的波动上，保持很多的头寸。对于持有多头的基金来说，基金的赎回动作累积起来，会造成很大的价格波动。在本节中所描述的机制使得统计套利陷入困境。

超强的毁灭效应

如果清算一个由普通股组成的、大型的统计套利投资组合，肯定会对其他类似的统计套利投资组合造成损失。组合越大，其影响也就越大。因为低价卖出多头，与高价买进空头，一定会对其他类似的组合产生影响。除非将清空头寸，把损害降到最低，否则损益的波动率，可能会对你造成很多的损害（包括你的胃口）。当毁灭发生时，由于价差交易的多头和空头方都会受

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

到不利影响，因此其影响非常强大。

在 1994 年 11 月，据报道，Kidder Peabody（一家证券公司）清算了一个超过 10 亿美元的匹配交易投资组合。美国长期资本管理公司，除了广为人知具有高度杠杆的利息交易头寸之外，还拥有一个在 1998 年 8 月被清算的匹配交易投资组合，造成了巨大的损失，因而威胁了（并最终破坏了）其偿债能力。

8.5 《公平披露条例》

《公平披露条例》是在 1999 年 12 月 20 日，由美国证券交易委员会提出的，而且马上就产生了现实影响。在官方采纳这个条例的 10 个月之前，一些现在声名狼藉的华尔街分析师都还肆无忌惮地从事一些过分的行为。例如，他们向客户推荐一只股票，但私下却并不看好它。或者他们会将一个负面评级改成正面的评级，以完成交易，然后再恢复负面评级！

1999 年，那些被《公平披露条例》认定为违法的行为，对统计套利投资组合产生了负面的影响。一般在消息正式公告的前几天，很明显某些人获得异常的收益。一般来说，一些分析师会从 CEO 或 CFO 那儿得到一些暗示。分析师与部分客户便在消息被公布之前，做出一些交易。如果是好消息，这些人买进的压力会导致股票的价格升高，统计模型会对这个相对强的股票发出做空的信号。几天之后，当这个消息被公布时，大众的狂热会再次把股价推高，造成统计模型做空的头寸损失惨重。反过来说，如果是不好的消息，导致同样的损失，而使得统计策略在短期上损失惨重。

这种行为模式在当时相当普遍，以至于许多市场参与者对此很有意见。美国证券交易委员会听到了人们的不满，并迅速采取了行动。一夜之间，那些不良行为便消失了。

带有特权性的消息被送到分析师的手中，并不是 1999 年的一个新现象。滥用特权是当时出现的新情况，或者刚才描述的事件来看是这样的。如果滥用特权早就存在，那就是我们没有注意到而已。还有一个有趣的事情，就是

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

分析师的影响力问题。在《公平披露条例》公布之后，有许多之前能准确预测公司绩效的星级分析师的表现变平庸了。

8.6 在亏损期间的相关性

当一个持续获利的投资组合出现亏损时，投资者会哀叹：

“你的结果，与其他管理者的结果‘高度的’相关。”这暗示，由于所做的交易类似，即便是通过不同的股票选择方式，不同的交易识别（预测模型）方式，并且使用不同的交易特征，比如不同的持有期，可是最后并没有产生差异化效果。这些差异化的方式都是假的吗？绩效的相关性都是一致的吗？

将两个有区别的但同样以大量的股票为基础的股票投资组合，通过一个反转模型来进行交易，那么在亏损期间，二者的表现会出现高度的一致。如果正常的市场行为（正常的时期是指价格移动模式受到投资者的活动的影响）被一个事件破坏，像是全球次贷危机以及战争之类的事件，事件对整个股价有一个明显的影响：最近的相对强弱趋势会使反应情况升级。在事件的影响期间（大家都认为“事件”是“坏消息”的同义词），总是会产生低价出售的结果。与那些较强的股票（至少是不弱的股票）相比，会先卖出那些较弱的股票，其影响范围比较大。这与基金赎回的预期效应正好相反（见 8.4 节）。这对价差交易来说，其含义很明显：亏损。不管管理者的策略、交易的领域以及在那段期间所采用的交易头寸有何差异，所有的价差反转交易都有一个共同的特性，就是在一个时间点，被判断为相对弱的股票会被做多，相对强的股票则会被做空。那么市场必然出现亏损。

不同的管理者，亏损的大小、亏损持续的时间以及恢复的时间，都会不同。个别的反转模型和管理者的风险决策，会对它们产生强烈的影响。

任何在经济上、政治上或其他导致投资者恐惧的事件，都会慢慢灌输卖出的概念。这对所有以大量的股票为基础的价差投资组合来说，采用平均反转的方式会有明显的影响。很显然的，绩效表现会转为亏损，不同的管理者的相关性还是很高的。有趣的是，相关系数可能不会很高。对于不同的投资

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

组合来说，在亏损期间的回报大小可能不会相同。在恐慌之下卖出的理论中，不会出现有序行动的建议，不会要求在市场之间、公司股本之间进行平均的分配，或有其他的干预行动。一般来说，坏消息出现时，总出现手忙脚乱的情形。因此，管理者无法避免经历异常亏损的期间，而在亏损期间的实际回报，与获利期间回报的相关性，可能是正的、负的，也有可能是零。

经历了一段亏损期后，总是会迎来盈利期。按照前面的说法，市场崩溃的情况总会得到缓解，在价差反转策略的亏损或盈利期间，会创造出相似的模式。

在价差交易出现盈利的期间，不同策略的表现又会怎样呢？这是预测，不同的策略带来的回报也不同。回报与模型的精确程度是密切相关的。很难找出一种统一的方式，使其与“恐惧”在程度上相似但效果相反，作为短、中、长期现象的参考。可能繁荣是最接近于这种方式，因为它能够随机地产生一些反转的机会。但是同“恐惧”相比，“繁荣”就没那么具体了。它不像恐惧那样，能够使人们共同行动。同亏损的反转交易的期间相比，不同的管理者的投资结果会不同，而且具有很大的差异。

图 8-3 举例说明了这个典型的情况。从整体上说，基金 A 与基金 B 的回报呈现出正相关性，相关系数为 0.4。这个相关性的结果表现在两个象限中，并且大多数交易结果在这两个基金双赢或双输（正 - 正与负 - 负）的象限内。在那两个象限之中，交易策略的绩效非好即坏，它们的相关性却都是负的：在亏损期间，相关系数是 -0.19，在盈利期间，则是 -0.22。这似乎是矛盾的，总相关系数是正的，但在两个象限中，相关系数却都是负的，这恰好证明了相关性的谬论。请注意，在亏损期间，其回报之间的相关性（为 0.19）实际是比较低的，比在盈利期间还要低（为 0.22），这与先前描述的正好相反。这个示例说明了，文中论述的共同主题，那就是尽管我们可以识别出各种各样的模式，并对其特点（总体的或平均的）进行描述，但还是存在一些差异，这些差异我们还没能掌握，却必须要面对。还要注意的是，在亏损的期间，只有 10 个数据点，仅占盈利期间的 1/4。因此，相关性的估计并不准确（以统计学的说法，自由度过低，或者数据太少。估计变量

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

之间的关系，10 个数据太少了）。

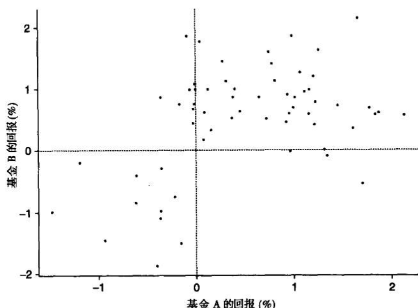


图 8-3 基金 A 与基金 B 的月回报，说明了相关性的错误

无论采用哪一种价差反转策略，都必然会遇到亏损期，这不会令人感到惊奇。应当理解在两个特别的年度中，相关性会受到市场的影响。理解如何得出结论是很重要的。考虑亏损与市场之间的巧合关系，还不如寻找控制亏损的方法。应该将注意力放在对于价差反转驱动力量上。我们应该考虑那些驱动的力量，是否会再度出现，何时才会变得足够强大，为管理者产生可以系统地获利的机会（见第 11 章）。在这里，辨别出未来可能的赢家与输家是有可能的。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net



第 9 章 多重困难

某种做法，虽然从长远角度来看，可能会带来致命的后果，但只要短期内具有一定效果，就会被采用——而这往往导致了灭亡。

——现代综合进化论奠基人　杜布赞斯基

9.1　导论

在经过了差不多 20 年的高盈利期后，从 2000 年年初开始，很多统计套利管理者的回报，一下子就降到零，有的甚至更糟。一些管理者在接下来的两年中，还有很高的回报，但是到了 2002 年年初，这也消失了。在统计套利的历史中，2000 年是盈利和亏损的分水岭，市场进入了一个高频率反转的动态变化时期。因这种普遍的效应，带来了更加显著的影响，很多人认为 2002 年是绩效的分水岭，标志着统计套利的灭亡，不再能产生绝对回报。尽管一些观察员认为这只是暂时性的结构问题，等到条件成熟时，统计套利将东山再起，但其分析不是那么充分，投资者的耐心更是消失了，导致投资人纷纷撤资，基金管理者连工资都支付不起。

2005 年年底是统计套利最可怕的时期：作为一种投资产品，它成为了滞销品。到了 2006 年，基金的绩效出现复苏的希望，同时出现了各种各样的说法，试图为当年溃败性的表现进行辩解，但其中许多解释都经不起考

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

验。在暂时中断市场的连续性和可预测的证券价格动态变化之后，人们希望统计套利能够带来一个新的获利时代。我们已经进入新时期有两年的时间了。

在本章中，我们将检查几种宣称能解释统计套利回报下降的原因。虽然每种说法都对统计套利的绩效有负面影响，但它们不可能会造成回报下降超过30%。或许可以说，这些说法确实对统计套利的绩效有显著影响，但不是决定因素。接下来，我们将观点展开，考虑美国经济与金融市场的主要发展，并描述其对统计套利的冲击程度，尽管很多都只是暂时的。我们将讨论关于2003年的情况。

在第10章，将会继续讨论这个主题，寻找绩效表现下降的原因，以及复兴的来源。大型经纪商在交易技术发展上，会长期影响统计套利的策略，也会产生负面的影响效应，但对大型市场参与者来说，如果能更多地使用这种工具，就会发现很多新的机遇。

9.2 十进制

"股价的单位由25美分降到2美分，使得统计套利的优势消失。"

很多策略都以"统计套利"冠名，但它们之间有一些很重要的区别有助于理解，为什么它们近几年的绩效不同，对未来的预期也不相同。从2000年中期开始，高频率统计套利策略投资的参与者的回报很凄惨。事实上，很多人在2000~2001年期间是亏损的。在此期间内，买卖双方之间的价差也在不断地缩小，价差单位从1/8美元，下降到1/16美元。这对于那些人的回报产生了巨大的负面影响。此外，场内的自营商开始联合起来，形成了5个主要的机构和两个比较小的机构，导致每天的价格反转交易被自营商内部消化。在研究预算、计算机、人力资源等方面，自营商比大多数的统计套利基金管理者更有优势，而且能看到订单的流量，又使他们具有不公平的优势，因此会发生这样的情况，也没什么好惊讶的（见第10章）。

统计套利的优势并没有消失。不过高频率的机会不会出现在公共领域，

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

只能为少数有特权的人把持。策略头寸在持有的期间不是几天的时间，而是几周或几个月以上，交易策略也就不会轻易地受到报价单位十进制的影响。对于解释本章一开始那段描述的 2000 ~ 2002 年绩效所产生的差异，模型的动态特性有很重大的意义。对已采用高频策略的反转过程来说，买卖方报价之间的价差单位变小，不会影响长期的交易利润。一些绩效衰退的现象，可能源于比较差的执行情况，而不是报价单位十进制。如果想要弄得更清楚，我们可以来看一些示例。

看一只平均股价为每股 40 美元的股票。假设交易策略的目标是每年收益 12%，也就是平均每月 1%。以两个月的持有时间来看，一般来说，一个反转交易能得到 2% 的回报，也即是 80 美分。买卖方报价之间的价差一直是以 25 美分为单位，由于十进制而降低到几美分，这导致的损失，最多达到获利期望值的 1/3。因此，年回报率将会从 12% 下降到 8%。这是最差的情况。如果运用这种交易战术，或许还有改善的可能性。但对于一个高频率的交易策略来说，不存在这个问题。报价单位的十进制，对于较长期交易策略的实际影响是很小的。

当然，不会只谈到报价单位的十进制就结束，本章随后的内容会涉及更多方面。在报价单位十进制的过程中，长期的统计套利交易策略似乎具有更显著的优势，使 2000 年与 2001 年间统计套利策略能够继续生存下去。当到了 2002 年和 2003 年，市场发生结构变化时，这种交易策略就毫无作用了。

9.2.1 欧洲的经验

在十余年前的欧洲市场，报价单位就采用十进制，同时高频率统计套利的策略还是发展得很顺利。可以得到这样的结论，报价单位采用十进制并没有阻碍策略的获利。在绩效表现上出现的问题，是由市场结构的变化导致的。这导致了时间在动态上的改变，从而将使得统计套利从业者建立预测模型的模式失灵。欧洲市场完全采用电子化的方式，比较接近纳斯达克，而不是纽约证券交易所。然而，2003 ~ 2004 年期间，统计套利在上述所有的市

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

场中，都未能取得回报。在这三个市场中可以用各自不同的原因来解释绩效表现，但在这三个市场中，存在一个共同的影响因素。那这个共同因素又是什么呢？2003 年上半年可以认为是受到了伊拉克战争的影响。但之后的六个月呢？2004 年呢？又是被什么影响呢？虽然在当时，美国的超额预算与贸易赤字，使许多人恐慌（主要是评论员和经济学家），但对经济新闻的报道，还是使人们逐步从悲观态度转变为乐观态度。正是这种变化以及投资者相应的变化，导致基金管理者没有因统计套利而获利。在每个市场里，特定的影响因素也是存在的。

9.2.2 恶魔在鼓吹

刚才谈到了，除了一些高频率的策略之外，报价单位十进制并不算是导致统计套利回报减少的重要原因。让我们来看看，这样的变化是如何带来不利影响的。

报价单位十进制使流动性发生了直接的改变，这中间有许多饶有趣味的故事：流动性的改变，与每天的价格行为以及统计套利的交易成交量有什么关系呢？初步的证据显示，交易策略回报的下降，是由于市场价格模式的改变。然而又是如何改变的呢？是以何种方式改变的呢？在我们所观察到的改变，与报价单位十进制以及统计套利绩效表现之间，有没有一个合乎逻辑的关联呢？暂时不予考虑十进制对一天之内定价机制的改变的合理性，让我们假设这种改变的主张是正确的。这对于每天的价格行为来说有什么意义呢？

在其他因素保持不变的条件下，由于市场价格增量的大小不同，从早上 9:30 到下午 16:00 的交易会导致每日价格模式不同。有没有一个程序能对此加以说明呢？如果没有，那么除了交易策略能获知收盘时买卖方价差之外，用每日收盘价格作为系统交易模型的计算参数不会影响计算的结果。事实上有很多采用收盘价格的模型，对 2003 ~ 2004 年期间的数据进行模拟，只能获得可怜的回报。由于报价单位采用十进制，以美分作为报价单位，每日的价格波动出现了明显的、公认的、结构上的变化，导致收盘价格在时间上产生结构性的变化。或者与报价单位十进制无关的其他因素，才导致了这

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

种模拟结果。

如果从相反的角度观察，那么可以从报价单位十进制和系统交易策略回报之间得到一个貌似正确的结论：假如交易都带来利润，但是在同一天的交易的回报为零，那么价格随交易而变动，导致丧失了获利的机会。现在的证据既不支持也不反对此假设。在对统计套利绩效表现下降进行解释时，不能说报价单位十进制对此毫无作用，应该说它起到了重要的作用。

假设统计套利在历史上的绩效表现，来源于系统性的获取消费者盈余(consumer surplus)。当价差突然发生跳跃，超过了进行交易的门坎，市场参与者用高于这个“超额”的价格成交。报价单位十进制将这种突然的跳跃降到零的程度，根据以往的经验，我们就可以合理地假设价格会提高，那么交易的回报期望值也几乎降为零。这个方案颇具有吸引力，直到人们慢慢认识到，这个方案和关于买卖价差的结论是相同的。价格跳跃导致的消费者盈余，就是买卖价差（跳跃）变大的结果。除非价格变大是策略回报的一个主要部分，否则这个结论就毫无意义了。

9.3 统计套利结束了

“统计套利两年来都无法带来利润，它的优势已经完全结束了。”

在2004年，这样的说法广为流传。但有什么证据能够证明这种说法呢?除了最近的历史观察结果之外，没有其他的证据，我们只能认为统计套利的绩效是这种说法的直接证据。如果这样，追溯更久一点的数据，就能将这种论述驳倒。从1998年年中到1999年年末的这18个月中，这个策略的回报几乎为零（以因子为基础的模型比其他模型要强一点），然而在随后的两年里，它获得创纪录的高回报。

在一定的时期，由于绩效平平就推断运用股价行为模式获得回报的交易方式已经消失了。同样，不可能只通过检查一些数字就认定，只要看到糟糕的情况过去了就必定会获得高回报，或是任何特定模式的回报。要想知道什么才是可能的，人们就必须了解股价波动的特点、交易策略的不足之处以及

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

如何运用交易策略。要想进一步了解详情，人们就需要对这个世界各种状态做出描述，并进行预测（见第11章）。当然，以现在的数据看，也能驳倒这个观点。因为统计套利获得了相当不错的回报。

9.4 竞争

“竞争消灭了统计套利这种游戏。”

这是一种倨傲的评论方式。它将统计套利当做一种游戏，在实践中有人把它当成轮盘赌，我们把这种观点放到一边。连赢是可能的，但是连输时，赌徒们就开始恐慌，采取轻率的行动，然后到绝望，最后从市场中离去。那些了解交易策略背后驱动力量以及其执行方式的人，通过制定交易纪律与控制方法，使自己能够在市场的冲击中幸存下来。

竞争使风险套利作为投资策略消失了吗？是的！2002～2005年，统计套利机会的消失，并不是因为统计套利人数的增加，也不是因为统计套利管理的资产增加，这两个因素都使统计套利的回报下降。真正的原因是因为经济结构发生了变化。在2005年年末，人们预测并购事件的上升，使并购套利复兴，而且进行了几个月的经济报道。从事统计套利的人员增加，活动量也变大，影响了套利的回报。一般来说，获利也会变小。如果那些比较优秀、经验丰富的管理者发现并应用了第11章所讲解的交易模式，那他们将获得不错的收益。一些依赖传统方法的新手，会有一段艰难的时期。运气也是至关重要的。

由于恐怖袭击、战争以及一系列公司违法行为导致经济结构发生变化时，并购与统计套利之间到底有什么差别，暂时中断了一些人的事业，同时终结另外一些人的事业呢？（现在知道是一个错误的判断）非常重要的一点是，必须知道回报的来源，并且有什么样的限制条件阻止产生回报。当提到并购套利的时候，“交易流”这个不可思议的单词就出现在投资者的脑中。这种深入的理解，提供一个深入思考的契入点。但经济环境得到改善（没有进行定义，也就是自身的理解），投资者愿意承担风险。兼并与收购就会出

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

现。在中场休息之后，游戏也重新开始。统计套利的驱动力量没有这样的标识（尽管"波动率"也是一个希望，它也确实可以）。我们并不理解统计套利的运行原理，与宏观经济发展也没有任何联系，也没有可供观察的指标。人们还必须深入地思考。那是很困难的。因此，存在很多的不确定性，混淆了视听，不可避免地出现了"统计套利已经死了"这种保守说法。我想知道，统计套利将复兴吗？

要严肃地关注有关竞争方面的争论。尽管没有公开的数据，说明对冲基金和投资银行在自营交易中，系统性地投人股票交易中的资本量，但从清算经纪人和投资者评论，以及媒体批露出的信息中，可以推断出，在2000年以前，基金的数量以及投入的资金量大幅增加。以这个观察结果做为证据来支持竞争假设，马上就会遭到反驳，因为资产与管理者的数量在这20多年来一直持续增加，资产类别与统计套利的绩效之间的联系有限。随着资产或是管理者的增加，统计套利的绩效并没有下降。这个假设仅仅说明在那两年中，表现优异的绩效出现了中断。为了要解决"好像正确，却没有充足的证据"这一问题，这个假设还需要说明竞争因素为什么只在最近才变得那么明显。

从2000~2002年期间，市场行情急剧下跌，先前对统计套利和对冲基金都不感兴趣的投资者，增加了很多投资方式做为备选方案。因此，可以这样说，统计套利投资在2002年呈现了一个阶梯式的变化（增加）。

但这并不是在2002年初发生的，不是吗？

除了投资的数量和管理者数量外，还有什么其他的证据，来支持或否定关于竞争因素的假设呢？到目前为止，这些肤浅的结论，只是让我们更加关注凄惨的绩效，这甚至不能说是一个结论。它只是观测结果与原因之间的一个巧合，并没有说明其中的运作机制。最简单的情形就是，"很多管理者竞争，用相同的价格交易相同的股票"。仅仅通过流动性，市场的参与者就能发现这种类似的情况，例如，已完成的交易会影响未完成的交易。还没有一个广泛接受的证据来证明这个解释。

假如在交易期间，竞争因素使"消费者盈余"与价格波动（从历史上

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

看，这是统计套利利润的来源）消失，那么应该可以在每日收盘价格中看到这种影响。事实上，很多交易都获得消费者盈余带来的收益，这不符合前面的结论。统计套利的机会大幅缩减，与波动率直接相关。统计套利管理者之间的激烈竞争，才能导致波动率在一定范围内降低。现在依然遗留了一个重要的问题：对于那些被识别出来的交易机会，为什么总回报会降至零呢？

在证券交易中，竞争在不断地增强。尽管统计套利的绩效也确实出现了问题，没有证据可以证明随着市场冲击力的增加，激烈的竞争导致统计套利绩效不断下降。

9.5 机构投资人

“养老基金与共同基金在交易中，变得更有效率了。”

在三年下跌趋势的影响下，机构投资人努力降低成本，因而这种回报也消失了。降低交易成本，成为那些机构投资人的主要目标。大宗交易过去确实带来了许多反转的机会，但由于那些交易者变聪明了，导致那些机会也就消失了。经常可以听到“Fideliy 公司采用交易量加权平均价格（VWAP）已经好几年了”这类说法。我们要再次注意这一点，这些变化并不是一夜之间就发生的。在某种程度上，这些变化可以当成是统计套利绩效表现下降的一个原因，它与其余的原因混合在一起不能分开。评估这些变化所造成的具体影响是一件不可能完成的任务。的确，机构投资人是第 10 章描述的交易工具的主要使用者，他们造成的实质性影响也是毋庸置疑的，这些影响对统计套利来说是负面的。

9.6 波动率是关键因素

“市场波动率不停地变化，对统计套利有利吗？”

从 2002 年年初开始，人们就试图解释为什么统计套利策略回报不尽如人意。尽管 2001 年对很多人来说是不错的一年，但对有些管理者而言是匮

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

乏的一年。2002 年之后，(几乎)所有的管理者的绩效都很差（见图 9-1)。市场行情在不断下滑，波动率被当成统计套利绩效很差一种原因。后来以波动率为基础的解释，很快变成了唯一的一个主要的原因。这种观点似乎为绩效表现很差提供了一个“银弹”，评论员们几乎都忘记了它只是个假设，把它当成了唯一的原因！本章及后续部分都是为了解释：很多“小”的因素共同导致了统计套利无法产生令人满意的回报。

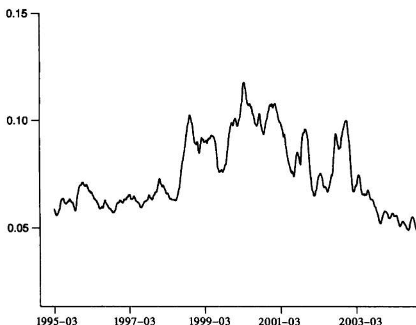


图 9-1 标准普尔 500 工业指数的平均的局部波动率

价差交易利用的是股票之间价格移动。股票之间的波动率对绩效来说，是至关重要的，股票之间的波动率虽然受到市场波动率的影响，但与市场波动率之间并不具有简单函数关系（见第 6 章)。股票之间的波动率通常与市场波动率正好相反。在 2003 年第三季度，股票之间的波动率降到了历史最低点，而市场波动率却在增加。在 2004 年，股票之间的波动率一直都在下降。

正因为股票之间的波动率与市场波动率并非简单相关，所以股票之间的波动率的水平，与交易策略的获利能力也不是简单相关。从整体上来说，价差模型只是利用波动率的一小部分；股票之间的波动率对于交易策略回报具

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

有很小的影响（除了在2004年波动率降到最低的情况，以及能够获取更多原始波动率和更精密的模型以外），对回报的波动率具有较大的影响。通过对比2003年第一季度与第三季度的市场条件和交易策略的绩效，充分说明这种影响力。股票之间的波动率在第三季度时，达到了最低水平，确是当年第一次获利。波动率在第一季度时提高了20%，绩效却是亏损。

在2003年9月，股票之间的波动率下降到了有记录以来的最低水平，然而反转机会却比比皆是。不过两周的时间，统计套利策略就产生了1%的回报。

不幸的是，在这段期间，共同基金公司在从事频繁的短期基金买卖被披露，这种短期买卖的行为违反了基金本身的规定，引发了规模达44亿美元的基金赎回活动，9月下旬，价格陷入了混乱之中。通过这个故事（在10月10日星期五的《金融时报》中详细介绍了其中的赎回细节）可以得出，10月几乎一定会发生更多的"崩溃性"事件。晨星（Morningstar，在10月9日星期四的《金融时报》中）建议投资者，减持或者不再持有Alliance Capital和美国银行（Bank of America）的共同基金，而那些基金管理者也忙于选择交易时机。

利率和波动率

由于利率极低，一年前（或者五年前、十年前）的美元价值，与今天美元价值在本质上是相同的。当利率较高，时间也就具有价值，"成长价值"这一概念也会有很大的不同。由于价值评估均衡化和一些判断因素不再有效，因此股票之间的波动率也缩小了。比较高的利率会增加股价的鉴别能力，增加反转机会的数量。从2004年年底开始，利率逐渐增加。美联储通过长时间的、未间断的方式将利率提高到了5%，因此从2006年开始，统计套利又再次带来了相当好的回报。

其实，从2004年开始，波动率就没有增加了。实际上，它甚至降到了有记录以来的最低水平。波动率还会升到历史的最高水平吗？绝对有可能。交易量加权平均价格上升，更多的竞争，以及第10章中所提到的交易工具，这些前面提到的变化，都暗示现在还进行交易，那就太傻了。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

9.7 关于时间维度的思考

前面的分析是在静态下进行的：从计算的结果来看，报价单位十进制或者竞争等因素都对任何的个别交易（平均看来）没有明显的影响（从保守的角度来看，还是会有影响的，但不会造成回报为零）。统计套利不只针对股票、成对的股票或者一堆的股票进行独立交易。它是与时间相关的一系列交易。交易的时间影响着交易策略的利润。在同一个时间点上同时间完成交易，是一种交易方式。前面论述，竞争导致管理者不能对识别出的机会加以利用。但竞争，报价单位十进制或者其他的因素都有可能会改变价格的时间结构，以及价格的进化过程，原本的那些模式，就无法再产生回报了吗？股价结构短期的变化，会不会使得系统性的交易模型变为噪声模型呢？如果这样，引起这种改变的原因能够被识别出来吗？报价单位十进制或竞争等因素，对此有影响吗？这些因素是活性剂，催化剂，或仅仅是巧合呢？这些因素还在发挥作用吗？如果还有其他的因素，那么他们是如何导致结构的变化的呢？这个过程是否结束了呢？是否建立一个新的稳定状态，还是会恢复到原来的状态呢？

再一次看一看讨论起始点，那些以前曾经可以得到优厚回报的统计套利策略，到2005年为止，它们已经有至少三年的时间，大都不能再产生一个适当的回报了。很多这样的交易策略是一年亏损或是多年亏损。在前面的分析中，考虑了各种假设条件，而策略交易头寸的绩效表现产生了挤出作用，无法再合理地解释其所观察到的交易模式以及所识别出来的机会。相同的背景变化（报价单位十进制，竞争的因素或者是其他同时存在但不为人所知的因素），是否可能造成了股价行为在“时间结构”发生变化，使模型的识别能力和预测力不复存在呢？如果我们把所有的模型信号都认定为噪声，那么根据系统性的交易策略，会出现什么结果？平均来说（也就是总计来说），这些的回报期望值必然为零。由于存在随机因素，以及交易执行的管理者的意见不一致，使得系统性的策略只能获得零回报，甚至是亏损（交易成本使

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

回报为零的原始交易，形成亏损）。

从表面上看，三年来，统计套利这一交易策略的回报基本不高。是假设符合观察情况，还是假设促成了这种结果出现，能够找到具体的证据吗？假如一个模型没有预测能力，那么对于一组已识别出来的交易来说，预期回报就应该与实际回报不相关。我们掌握的一个与基金有关的证据，很明白地否定了这个假设。对于这三年中大多数的交易来说，预期回报与实际回报之间呈现正的（从统计学上说是显著的）相关性。有很多的交易，都产生了正的回报，尽管平均来说，这些回报比前些年稍微低了一点。少数亏损的交易造成很大的损失。对于很多统计套利管理者来说，虽然其他的一些因素也会对这个结果产生影响，但这就是绩效问题的关键点。人们可以做出如下描述：交易信号一直存在（一般来说具有较高的胜算）；收益减少（与完整的交易相比，其回报率比较低）；动态的变化呈现不稳定状态（头寸持有时间很不稳定）；环境的噪声也变大（亏损和获利交易的比例以及持续反转的位置，都有较大的变化）。

我们可以运用波动函数的原型，也就是池塘中的波纹，或者是正弦曲线，进行举例说明。这个示例将有助于说明不同成分分别对交易绩效表现的贡献，以及变化的含义（见图9-2）。

首先，请记住噪声是有益的。假设观察值只是一个单纯的正弦波，再加上几个随机数，用 $y_t = \mu_t + \varepsilon_t$ 来表示，其中 $\varepsilon_t \sim [0, \sigma]$。一个比较大的噪声变化，产生如图9-3所示的序列，较小的噪声产生如图9-4所示的序列。

通过对信号（模型）以及噪声成分影响的分析，在图9-3产生回报可能更大。我们知道信号会往哪里走，而且我们也知道多大的噪声会取代信号。这导致将单一利用信号预测修订为同时利用信号和噪声预测。当模型的预测结果为零时，可以先不清仓，根据噪声的分布情况，等待出现更有利的结果。这个现象就是随机共振现象（见第3章）。可以在 *A* 点进入，在 *B* 点退出，而不是在 *C* 点。同“利用信号”这一模型相比，带来的绩效是不同的（错失交易，机会成本等）。通常，如果模型校验的很好，就可以获利。在

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

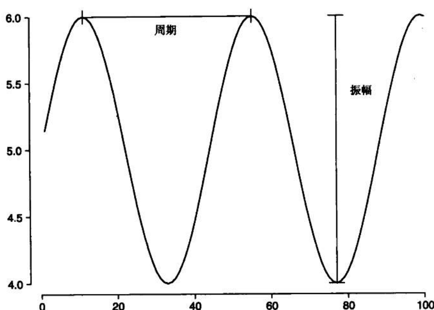


图 9-2 正弦波

图 9-4 中可以看得很清楚，依靠噪声所获得的利益，与机会成本（潜在交易产生的收益）完全不同，如图 9-4 所示。

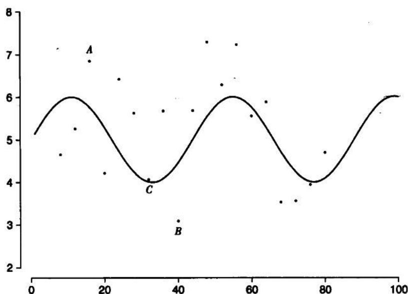


图 9-3 价差和正弦波信号

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

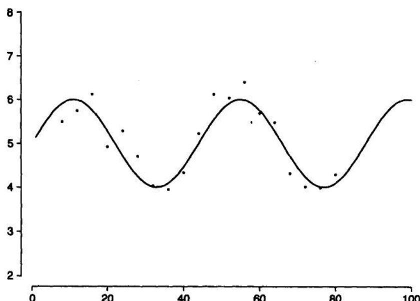


图 9-4 低波动率的价差和正弦波信号

在 2003～2004 年，很少有机构管理者进行交易。人们对这种无为状态进行了评论。"保持观望的态度"恰当地描述了当时的情形，大家很少会积极地进行判断决策，对于贫乏的经济政治新闻采取观望的态度，就更不必说那三年股票市场冷清的交易。不活跃的投资活动从不同方面影响了反转结构（尤其是由机构大量拥有，而且有专业分析师密切关注的股票）。首先这种情况使反转的步调慢了下来，增加了图 9-2 中正弦波函数的震荡周期。好像将图中的正弦波进行拉伸，而移动速度变慢。如果其他的条件没有改变，这样动态的移动，降低了交易策略的回报。现在的移动时间是原来的两倍，因此回报也只有原来的一半。

实际中的影响比这个原型还要大，回报甚至还不到原来的一半，因为其中还存在另外一些因素。其中最大的影响因素，就是模型设计者判断动态变化发生时间的能力，以及对模型进行修正的能力。如果在这个过程中利用数学原型，我们就可以立刻分析出变化（如正弦曲线的周期）的特点。但如果信号的周围存在大量的噪声数据，这个工作就会有很多困难。在很多情况下，变化会受到噪声的影响，噪声也会呈现不同的分布。变化只受到一个因

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

素的影响，或是造成变化的因素会按照顺序发生，而这种顺序会给我们的分析带来便利，这是不可能的。必须得花费一些时间来观察变化的结果，因为证据需要积累。由于在动态的结果中识别一个变化，在时间上会有所延迟，导致在利用系统信号时，会降低回报。

变化的过程增加了复杂性程度，而这也是回报下降的另一原因。在前面的讨论中都认为，从一个平衡点能瞬间变化到另一个平衡点。这种情况在系统的发展中，几乎是不可能出现的。更通常的情况是一种进化的过程。这个过程可能比较平滑，也不平滑，但如果信号的周围存在相当大的噪声，这其中的差异就没有任何意义。当把变化作为平衡点之间的路径时，这些变化是平滑的，还是一系列非连续而且大小不同的，结果都相同：当模型反映新的信号动态变化时，最终都导致回报进一步缩小。过度的波动和反弹的情况都使得波动率变大，因此，情况变得更加不确定。

如何应对价格结构及其他结构的变化，是模型设计者所面对的最困难的工作之一。不像物理过程，比如一个钟摆的运动，会受到摩擦与磨损的影响。在化学过程中，杂质则会影响化学作用。这些都有其相应的理论，却没有理论解释证券定价中的失真，及其对反转造成的影响。它不是一个机械的过程。只有在“正常”的情况下，也就是价格受到的干扰非常小，而且影响的时间非常有限，价格变化才表现出“机械性”。这样我们才有可能利用观察到一般过程，让投资者获得回报。一个好的模型应该能适应变化。例如股价波动率的变化，能自动地识别出来，而且能进行重新校验。自适应型的模型能够有效运转，有一个关键的因素，就是这个新的状态必须具有持续性。如果变化（或者波动率）没有持续性，先朝一个方向，接着又变为另一个方向，那么这种自适应型的模型，就会因为波形大小与方向变化的不连续性而失效。通过对市场的观察，一个模型设计者可以在模型中加入调节功能，根据实际的交易结果训练这种能力。难点是必须要选出一个模型校验的依据，决定什么时候不变，什么时候解除限制。这个问题要留给模型设计者，这份工作可以称得上是一门艺术了。在这方面，有的模型设计者极富天

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

分，大部分人是毫无希望。在利用模型的过程中，有许多人并没有了解当违反假设条件时，会对模型表现造成多大的影响。当情况会变差时，很多人无法采取补救措施。

这种补救行为很可能会以失败告终，再加上对于模型失效的了解（当价格模式偏离正常的轨道时，投资者如何做出一个连贯的解释），因此一个好的模型设计者，便会建立一个自动监视系统，设计出稳定的反馈机制，改善模型的绩效。

在了解了模型所利用的信号之后，设计一个监视方案时必须反复地思考这些问题："数据违反了模型的假设条件吗?""没有符合哪一个假设条件?"在一组条件下，模型的表现是否很差。假如观察到模型在某一个方向来回的反复调整，那么模型的设计者就需要对模型进行调整。

自适应型的模型靠反馈机制来进行调整的。如果股价波动率达到一个边际值时，那么就可以对这个模型进行重新校验。这个模型一直在某个范围之外（频率、大小、累积影响力）进行调整，对于监视程序来说，就是一个警告。一般来说，管理者无法做到前反馈机制，这是因为我们通常无法改变环境，从而影响股价发展的模式。据有关报道，某些大型基金会从事这种类型的活动，采取一些假的交易，先进行一些真实的交易，并将消息透露给其他的市场参与者。例如，买进 IBM 股票，等其他的交易者也买进 IBM 并使价格上升时，基金管理者就卖出它持有的股票，这是真正的意图。卖出的价格比虚假操作的价格要高出许多。管理者是断然不会承认这种行为的。技术的发展为这些秘密交易战术，提供了更多的机会，具体见第 10 章和第 11 章。

我们通过讨论机构资金管理者不积极进行交易的现象，观察这对价格的结构变化造成的影响，然后讨论如何监视并适应这样的变化。在剔除恐惧所造成的不稳定因素之后，发现这种不积极状态也会使（反转）机会减少。回头看图 9-2 中的原型，由于缺乏信号，振幅变得比较小（从最高点到最低点的大小）。两个相似的股票价格之间的价差之所以会变大，不是因为局部的价格波动，就是由于投资者追逐某个投资热点导致的。价格变动很小，好像

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

是价格被限制住了，这有可能是因为市场中的参与者都很谨慎（有些时候毫无生气）。环境（市场的主题、价格、行动）一般来说不会迅速地改变，但随着时间的流逝，机会就不复存在了。

当机构管理者发现回报降低（或者消失，甚至三年连续亏损），便会改变交易的战术。部分管理者试图降低他们的成本。传统的（大型）交易管理者掌控着大宗交易的执行方式，而管理者自身也参与交易，因而大大降低（号称如此）了经纪商的收费。这样的变化，①推翻了在 2002～2004 年，机构管理者交易不活跃的假设；②导致统计套利绩效很差。

我们没有证据来证明或推翻第一种说法。不过这种情况已经过去了，我们不去管它。对于第二种说法，证据显示了相反的情形。如果一个管理者负责一定数量的交易，采取更有效率的交易战术，或者是采取了一些更积极的行为，以降低市场冲击，那么针对那些被机构大量持有的股票，在统计套利中会产生何种影响？

在某种程度上，大宗交易是造成股票之间价差扩大的原因（为反转交易创造了进场点）。将大宗交易分成许多比较小的交易，就会减少反转机会，反转信号也会变少。如果有某个管理者将资金转移到药品类的股票上，采取了适当的行动，Pfizer（美国辉瑞制药有限公司）和 GlaXo（英国葛兰素史克公司）的股票价差只会有一个很小的变化。对市场影响力比较小，反转的机会也随之变小。可能带来利润的反转机会也会减少。由于市场价格波动还受到其他因素的影响，因素的影响并不是那么明显。平均反转数量缩小，会降低交易策略的回报。然而，交易战术上的变化会产生其他的影响，故现在做出结论还为时尚早。管理者的行为变得更加积极，使反转发生的频率增加。反转的频率增加，它的回报也会增加（假设有足够的资金提供反转机会）。管理者是否会减少降低价差，使反转机会变少呢？不过当看到一个价格有误时不去修正它，这也是很不合理的。你不可能面面俱到的。

有证据可以证明，机构管理者进行了数量众多但规模小的交易吗？在 2002～2004 年，反转发生的频率降低了，而并不是加速了。不过关于反转

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

机会多少的证据，就没那么明显。确实有一段期间，股票之间的波动率降到了最低点，例如2003年3月，当时大多数股票的移动都出奇的一致。但波动率的减少是受全球性证券的影响，这与资金管理者，几乎是毫无关系。

9.8 虚构情节中的真实示例

所有关于统计套利绩效变差的分析中，都包含了一个事实。每一个因素对于统计套利模型的回报，都会造成一定的负面影响。但是，把所有这些因素都加起来，所产生的影响也只有不到“正常回报”（指以2000年以前的回报）的30%。因此我们被迫去寻找一个更广泛的理论，用来解释统计套利绩效变差的原因。在前一节讨论到在时间的动态变化时，给出了一些暗示。现在，我们将会进一步阐述。

9.9 恶劣的行为

表9-1列出了在2002~2003这两年中发生的一系列事件。每个事件对金融市场的行动都产生重大的影响，超过了“一般的”变化。大多数事件都带有很大的负面影响，也都令人反感。

表9-1 事件列表

日期	事件
2001年12月	安然公司破产
2002年1月	会计丑闻，首席执行官与首席财政官渎职 华尔街研究：“谎言，该死的谎言，以及成为百万富翁的分析师”
2002年8月	注销公司账户
2002年10月	共同基金投资者的恐慌
2002年11月	SARS
2003年3月	伊拉克战争 股息税法修正案 纽约证交所/格拉索（Grasso）赔偿丑闻
2003年10月	共同基金短期交易丑闻
2003年12月	统计套利投资者逃跑

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

在2002年的头几个月，出现了一系列可怕的消息，这些消息与一些大公司领导人和华尔街有关，对大众的影响极大。这些事件把那些没有从恐怖袭击事件中恢复过来的大众推入更大的震惊之中。不出所料，金融市场与宏观经济活动一起，在股价的关系上出现了一个巨大的结构性变化，一件接着一件的事情产生加成放大的效应。人们没有休息，没有缓解，混乱在继续。

在2002年结束时，非典（SARS）又给了航空业和旅游业当头一棒。为了表示对亚洲禽流感的警觉，世界卫生组织（World Health Organization）在2004年下半年预测，在亚洲可能会爆发全球性的传染病，并可能会造成5 000万人丧生。

现在，这种恐慌全都消失无踪了，部分政治领导人关注这些紧急事件，就好像根本没有发生过一样。人们的反应与几年前对SARS的反应是一样的，难道是人们对于恐慌已经厌倦了吗?

除了SARS之外，世界人们也很关注美国在海湾地区不断增加兵力。美国会入侵伊拉克吗?到了2003年3月，美国的确入侵了。在“战争活动”的这三个星期，市场很难对部分投资者不理性的行为给出一个合理的解释。如果有爆炸的照片在电视上公开，或者是有报告分析了美军的问题，市场总是会下跌。而如果电视上公开的是美国军事装备移动的照片，或是巡航导弹像雨一样落在沙地上的画面，市场就会向上涨。这对于那些久经世故的投资高手来说也太复杂了。假如这一切都不是真实的，却影响了许多人的生活，这太可笑了。

到2003年的第四季度，纽约州总检察长斯皮策（Spitzer）揭露部分共同基金的一些不法行为。成熟的投资者更容易接受坏消息。整个产业没有发生大规模溃退。投资者很冷静地从那些可耻的基金中撤资，并迅速将钱转移到其他的共同基金中。华尔街爆出更多不法的活动时，人们的反应已经变得足够理性，没有产生过于震惊或是未经思考就行动的恐慌情形了。毫无疑问，三年来的首次上涨也发挥了重要作用。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

这一系列具有影响力的事件，每个都如此的可耻，是骇人听闻的事件。再加上全球性的、具有破坏性的事件（战争与疾病）带来的影响，导致金融市场在两年中都处于不稳定状态。

市场的崩溃将如何影响股价的反转过程呢？图9-5将前面提到的价差原型（见图9-2）扩展到市场崩溃期。有些投资者不遵守市场规则，将价差推到一般变动范围之外，但是恐慌结束之后，或者是恐慌的反应消失而且投资规则重新建立之后，价差就恢复到通常的波动模式。

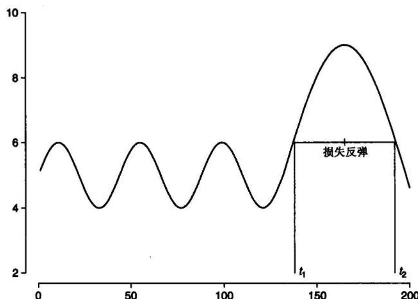


图9-5 带有暂时性干扰的正弦波

在 t_2 时点，模型对于价差的估计结果，不同于 t_1 时点（好像模型有一个不同的观点，但是稍微多考虑一点，就清楚地说明价差在时间维度上的变化。自适应型的模型包括了这种观点的进化过程）。通过模型向回看，信息（历史数据）效果会打折扣，影响未来模式的变化，例如将焦点放在均值上。往回看的时间比较短（即快速折扣），就会出现一个比较高的平均值（同时振幅和周期也都比较大），而且退出信号也会过早地出现，从而导致交易亏损。往回看的时间如果比较长（即比较慢的折扣），就会产生正常利

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

润（就像本示例），但是延长了持续的时间，回报也会降低。没有什么时候，能快速获得收益。因此，很显然，回报必然会下降（在没有进行干扰的情况下）。

下一节我们将从心理学的角度，对 2003 年进行特别的剖析。通过分析市场参与者的变化特点和原因，以及对证券价格发展的影响，进行一个严肃的分析（虽然是限定在一定范围之内）。它以事件描述为基础，目的是披露，美国在宏观经济、政治，以及金融市场上表现出来的变化。9.11 节将描述，这样的结构性变化会对计量模型产生什么影响。为了管理这样的变化，模型设计者能做些什么。9.8 节预示了将要讨论的内容。不能用简单的一句话将这些描述和分析说清楚。这个世界不是那么让人容易理解。统计套利在三年内，都没能带来回报，有很多原因，包括不同的策略本身和不同的时间点。2005 年发生的事情，肯定不同于前两年发生的事情。如果想只用一两个很简单的理由来说明失败的原因，最后就会像盲人摸象，没有任何作用。它可能使复杂的问题简单化，但也会造成误解。太过于简单的原因让我们忽略了假设条件对统计套利在现实中的应用影响。从 2006 年开始，统计套利的绩效开始复苏，在 2003 ~ 2005 年对统计套利的攻击的言论都消失了。第 11 章将论述，统计套利如何书写新的篇章。

9.10 对 2003 年的剖析

2003 年美国准备并打响了第二次海湾战争，投资者对是否持有头寸犹豫不决，因此 2003 年第一季度的交易量降低了。由于持续三周的战争，投资者有时迟疑不决，有时特别冲动，有时狂躁。在 2003 年的第四季度，我们看见存在一定的短期波动性，市场环境正在逐步改善：不确定性正在减少；投资信心和意愿正在增加；公司通过实施一些长期的计划来取代一些短期行为，再次开始投资；个人乐观地看待市场上的机会和市场的稳定性。最关键的是，从 2002 年下半年到 2003 年 5 月，那种没有明确说明却无处不在的恐慌慢慢地消失了。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

2003年夏天，美国结束了战争，开始战后的复苏。人们在假期可以进一步的反思，有充裕的时间进行剖析。在2003年夏天，（价差）波动率比以前还要低，部分是因为股票之间的波动率在夏天来临之前就已经处于很低的水平，部分是因为人们都很清楚大家都需要休息。同时，出现了两个关键的变化。

在伊拉克服役的军人每天都在付出生命的代价。而一般百姓对此越来越不关心，就像对待高峰时间的交通事故一样：虽然是不幸的事件，但这就是现实。人们形容美国经济，通常会采用“成长”、“稳定的就业率”等字眼。如同教科书中所说的那样，通货紧缩恢复了它传统的角色。很少有人关心政府赤字与其含义，直到真正发生赤字为止。这些广泛的变化，给金融市场中留下了深深的烙印。

对战争的焦躁和经济衰退，逐渐转变为对战争的疲惫，以及对经济潜力和机会的兴奋。2003年下半年，市场价格变动没有任何规则，这反映出投资者既热情又犹豫、急促，又稍微有点奇怪而且不稳定。

9.11 结构变化的真实情况

市场情况的追踪与预测模型产生的混合信号，揭示了变化过程的复杂性。从2003年3月开始，统计套利绩效进入了荒漠期，这些模型认为，趋势将从坏的方向转变为好的方向，同时也认为会有一个持续向坏方向发展的趋势。在过去十几年中，这种好似精神分裂的现象表明，市场结构是混乱的、尚未经处理的，而且正处于转型之中。如果人们想赋予这些指标新的活力，就必须改进这些指标，使它非常客观地寻找平衡点，增加人们的信心。

市场价格变动反映感知方面、价值评估方面，以及市场参与者行为方面的巨大变化，并据此进行调整市场价格的变化，这将会是一项很困难的工作。在设计的模型中，利用这些已被识别的价格行为模式，这是不可能的（如果利用的模式一旦消失）。在市场出现结构性变化的期间，这样的模型

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

通常都会“失灵”。一些比较好的模型也很难管理巨大、突然，并一直持续的变化。

统计套利模型并没有特别的保护机制，来抵御市场突变带来的冲击力。在这种情形下，反转策略绩效与货币市场的回报差不多。这些情况对于反转策略造成了很不利的影响（处于恐慌中时，缺乏前后一致的决策）。用心选择交易投资组合，并在识别到信号时，严格坚持应用模型，专注于分析风险，是非常重要的。对于风险控制来说，理解模型运转的过程，也就是为什么这个模型能起作用，是非常关键的。股票的反转并不会消失，模型能够系统地识别出交易机会，并在实践中加以利用。反转现象发生的环境改变了，也要改变反转现象识别方法。

对于统计套利来说，在环境发生变化时，不能采用什么有用的措施，只能看着模型亏损。许多模型，尤其是好的模型，它们的制作过程都花了很多心思，希望在市场价格行为发生变化以后，依然能够使用并且有较好的预测结果。但是，不管模型的设计者付出了多少努力，当市场结构变化真的发生时，运气还是最重要的。在发生了结构性变化后，重新设计模型以适应新的结构，避免损失，也是很有必要的。

9.12 总结

到目前为止，我们得出了这样的结论：在 2000 ~ 2002 年期间，统计套利所产生的回报，有 1/3 因市场的发展而消失，这样的变化是不可逆转的。在 2002 ~ 2003 年，美国经济巨大的变化，导致历史回报大幅减少。从 2004 年起，对统计套利看法就产生了分歧。破坏性的事件发生的频率降低了很多。结构性的变化对市场还有些持续性的影响，但其对统计套利没有什么影响了。在 2004 年时，一些与股票相关的事件还在不断发生，例如默克制药公司回收止痛药伟克适（Vioxx）的事件，以及纽约州总检察长埃利奥特·斯皮策调查美国保险经纪公司马什·麦克里安（Marsh McLennan）涉嫌诈欺客户的事件。非常低的波动率代表着非常高的相关性，这限制了反转现象的

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

出现。现在相关性已经降低了，短期反转的机会增加。投资者纷纷选择指数股票型基金，“每个人都好像变成了一个指数”。通过使用这些高端的交易工具，限制了波动率的大幅变动（见第 10 章）。同时，这个因果关系带来了一种新的系统性的股票价格模式，也导致统计套利的复兴，这将在第 11 章中论述。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net



第 10 章 黑匣子出现

知道事物发生的原因，他便是快乐的。

——古罗马诗人　维吉尔

10.1　导论

20 年前摩根士丹利创造了匹配交易这种商业交易方式。在 21 世纪早期，摩根士丹利创造出新的商业交易方式，又重新占据了统计套利的前沿位置。这种新的商业交易方式具有比较低的风险和可持续发展的潜力，在商业交易的历史中是推陈出新的典范，它系统性地摧毁了过时的匹配交易方式在市场中的机会。算法交易的概念诞生了。从机构投资者到对冲基金，大量的交易订单产生，其中大部分都是通过电子交易实现的，这给人们带来了许多赚钱的机会，甚至超出了经纪费用的预期。结合对自营交易（自营交易开始于最经典的匹配交易）的了解和知识，再加上对交易订单流量数据库的分析，摩根士丹利和其他经纪商，找到了新的交易工具，将交易订单的数量和交易日当天的时间作为预测市场影响的一个函数，合并到模型之中，然后根据股票每天特定的成交量，对模型进行调整。

经纪商知道，只要简单地运用上述交易方式，就能有巨大获利空间，而且自动交易并不会导致市场系统性地下跌，因此他们将这样的工具提供给客

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

户使用。这样的工具在很短的时间内，能做出巧妙的决策。随着狂热的统计套利者出现，他们诱导使用这种交易工具的客户最终毁灭，而这些客户渴望通过这种交易工具获得新的优势，降低交易成本，或改善在执行价格上的边际效应。在这种新的商业交易中，有一些天才建立了组织机构，而且统计套利交易者的交易订单流量实时提供了大量的数据，但对于那些挖掘数据的人来说，欲望是贪婪的、无止境的，是很难得到满足的。

从这些挖掘到的数据中，我们可以通过特定的股票、一周中特定的某一天、一天中特定的时间点，以及当天交易的成交量，创建特定的交易模式。仅仅用一种可测量性的方式，来预测在某个固定的期间，有多少订单会以目前的价格交易特定数量的股份，这对于交易者来说，是极好的发展方向。对此，对冲基金已经试了许多年，他们利用的数据比经纪商要少很多，所以不能与经纪商取得的成绩进行比较。不管怎么说，这种优势已经消失。

如果应用符合逻辑的模型来分析交易订单流量的数据，并及时提供数据，就可以产生第一代的模型，让交易者对经常遇到的、急需解决的问题，得到数量化的答案：

- 如果在接下来的半小时内，买进 XYZ 公司 x 份股票，需要支付多少钱？
- 如果在这个交易日剩余的时间中，采用等待的方式来交易，需要支付多少钱？
- 如果在一个小时内，将市场影响控制在 k 美分之内，能以什么价格卖掉 XYZ 股票？

这些不需要宣传的好工具，能够产生出一系列交易机会，并且这些机会具有自我繁殖性。如果交易者应用这些技术，就会产生新的交易订单流量信息，而卖家就会将这些数据收集起来。可以调查下面两种不同类型的交易者，一种交易者没有耐心，属于“付钱就办完”类型，另一种交易者比较放松，属于“等等和看看”类型。从交易订单的流量、交易工具的配置以及交易者自己“提交或取消”的修改纪录中，可以自动为交易者建立模型。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

这样的模型可以估计出客户将会为一个交易支付多少钱，同时可以估计出在市场冲击力较小的情况下，需要花多长时间完成该交易。这些可能性让研究人员兴奋地尖叫。这些尖叫声是20年前上一代的交易者发现匹配交易机会时所发出尖叫声的回声和增强。

所有的这些机会自身都能带来利润，却没有资本的要求。自营交易的风险消除了，而这个“新的”交易变得具有无限的机会。

摩根士丹利当然也有其他的竞争者。高盛、瑞士信贷第一波士顿银行、雷曼兄弟、美国银行以及其他公司，都开发出了算法交易并将之市场化。

10.2 对交易成交量期望值和市场冲击力进行的模型化

先从数据挖掘开始。我们可以取得什么数据？哪些数据与回答“多少钱”有关？假设拥有XYZ股票10年来每天成交量的历史数据，总共约有2 500天的交易数据。第一件事情就是检查每天的累计成交量。每个股票每一天的交易模式都是独特的。如果你喜欢，可以把它称为脚印。假如使用“一双鞋适用所有的脚”这种方法，用一般哺乳类动物的脚印来预测大象的脚印虽然可行，但是肯定会出现不必要的错误（称为噪声或误差变量）。如果采用眼镜蛇的脚印（试着去描述吧）进行预测，结果将更糟。这样你就能看出问题在哪儿。

在数据分析与建立模型时，少量地应用数据的特点就可以解决这个问题。不管计算机处理多少个模型的变量，都不会存在问题。应该注意的是，如果过度使用了一些不必要的数据特点，同样也会导致预测的方差变大，因为有限的数据资源不可能推导出一个无限可分的信息库。数据被分割的越细，每一个数据段所带的信息量就越少。如果两个或多个数据段在本质上是相同的（按调查的目的来说），那么这些数据最好集中管理。此外，模型测试如果采用了太多的数据，就可能找到一个看起来很好，但实际上却是伪造的适配模型。在应用统计分析时，这些都是非常重要的，但同时也是经常被忽略的细节。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

一开始的时候，我们会抱着这种想法来分析数据，希望识别出某个交易日内交易量的成交模式。如何描述这种模式？由于 10 年前的模式与今天的模式是不相同的，用 10 年前的模式是不明智的。必须记住，由于技术和市场的发展，在造成市场发生戏剧性的变化之前，原来的匹配交易运用的反转模式，在经济上具有可利用性，并且存在了十几年的时间。通过检查 10 年前、5 年前以及今年的每日累积交易成交量的图形，你会注意到所有的图形（曲线）的看起来都很相似，同时差异也很明显——将近期的模式与早期的模式进行比较，发现成交量累积速度在每天开盘和收盘的时候比较快。在这种情形下，不能仅仅简单地汇总所有的数据，或者是估计一条平均的曲线。

先观察最近 3 个月的每日成交量模式，总共有 60 张图形。然后，看一组 10 年前横跨 3 个月的数据。你会注意到，基本的形状都是重复出现。但是成交量存在很大的差异：现在股票的成交量比 10 年前大得多。用每日总成交量的累积百分比，描述这些图形。所有的图形都会落在 0 ~ 100 的区间。这样做了之后，模式的波动性就会变小。所以，通过分析某一天的交易，可以知道这一天的成交量是高还是低。

我们如何运用这种观点呢？其中一个方式就是用曲线来表示成交量（一天中交易成交量的累积百分比）。用这种方式，可以计算出市场在某个特定时间点的成交量占全天总成交量的比例。换句话说，可以解答像“到下午 14 点，成交量会是多少”这样的问题。有很多数学函数的形状都是 S 形：由于概率分布在这里正是检验的对象，自然而然地会采用概率分布的累积密度函数。在统计模型的构建过程中，可以使用一种很方便的函数形式（我们至今还没有提到过的），就是对数函数。

先挑选出一个函数，将它应用到数据中。对于“到下午 14 点，成交量会是多少”这样的问题，你就可以得到一个量化的、与特定股票相关的答案。一般是对的。

某只股票在上午 11：30 时，已经累计成交 400 万股，那么今天的成交量应该很大。到下午 14 点，还会有多少股票成交呢？根据评估的模型，在

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

上午 11：30 之前，交易量占全天成交量的 30%，到下午两点的时候，交易量则会达到 40%。据此你就可以很容易地算出，接下来的 90 分钟，应该会有 130 万股票成交。如果你期望的交易数量为 100 000 股，你不需要额外支付太多的成本去实现这个目标，对吧？

在上面的分析中，只考虑到了交易的成交量，而没有检查交易中与价格有关的信息。因此，我们有必要进行一些纠正。以 XYZ 股票 60 天的交易数据来说，有很多买进与卖出交易，交易量从 100 股到 100 000 股。所有交易订单的成交信息都被记录下来。根据交易订单价格（或者是交易订单下单时的市场价格），画出交易订单大小与价格变化的图形，平均的成交价格会表现出一个很确定的关系（同时也有很大的波动）。对于价格的变动，根据当天的股票总成交量，进行比例上的调整，有些变动又一次神奇地消失了。交易订单如果没有超过“显著的临界值”，在成交量比较大的时候，带来的影响也会比较小。

用一个数学曲线，或者是统计模型，来分析交易订单大小（市场冲击力的数据），就可以回答下面的问题：买进 10 000 股的 XYZ，我需要花多少钱？请注意，买进与卖出可能会有不同的反应，这两种反应可能会与该股票当天是上涨或下跌有关。将原始的（60 天）数据分组，分别针对股价上涨与下跌的日子进行分析，就可以说明这种情况。更正式的方法是，针对股价上涨与下跌，设计一个能同时包含两种情况的统计模型，这个模型应该包含一个指示变量，验证估计系数。如果能够知道不确定的程度，并合理地计算出独立性与其他必要条件（不需要大量的工作），我们就可以验证这种统计模型的有效性。针对上涨/下跌的数据，或者是两种情况结合起来的数据，我们都能建立预测模型。预测结果之间差异，有实际意义吗？这个差异是什么呢？

对不同的数据采用不同的模型，会导致我们无法估计相互之间的影响（不同的成交量，上涨/下跌的日子，买/卖等）。如果想要理解模型结果，人们总会遗漏一些重要的问题，因为相互之间影响非常精巧，而且使得因素

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

的分析方式不能充分发挥效果。如果有人想要得到一个合理的预测，但感觉遗漏了重要的部分（如果有相互之间影响的话），那么这个部分在实际中（依赖相互之间影响）也就不重要了。

交易的时间点对市场影响也是很重要的——回想一下，前面对一天以内交易累计成交量模式的分析。如果选择的是一天之中交易速度比较“缓慢”，或是交易比较少的时间点，而价格波动又被限定在一定的范围内，需要更多的耐心，或者是放宽价格波动的范围，才能达成一笔交易。在前面的讨论中，对交易订单大小——市场冲击力的分析，没有考虑时间点的问题。很显然，我们本来可以考虑这个因素，最明显的方法就是，将一天中交易比较缓慢与比较快的部分进行细分（或者采用更简单的方式，比如，以每半小时为单位进行细分），然后针对每一个部分，分别估计相应的模型。虽然还有许多更复杂的技术可以用到统计建模与分析中但相对简单的细分方式在这里可以作为交易机会与方法的一种典型的示例（还有更复杂的示例，包括通过将时间细分的平滑函数，得出正式的模型化参数，以及采用回归树等分类程序识别分组的分类程序）。

10.3 动态更新

如果观察 10 年前与最近的每日成交量基本模式，就会发现交易模式已经发生改变。为了应对管理实践中的变化，当根据数据的估计得到了一个预测模型时，马上就会遇到如何对模型进行调整更新这样的问题。一种做法是采用最近的数据，建立当前要使用的模型。我们假设“最近”数据就是指 60 天之内的数据。随之而来的还有另外一个问题：这个模型什么时候进行调整呢？我们又要面对形态变化、进化比率以及动态更新方法等问题。这些问题同第 2 章的反转模型是密切相关的，其实与这里的情形也不会有很大差异。人们可以选择采用 60 天的周期，每天重新估计模型之间的关系。拿最近交易日的模式与最近的分布模式，或之前的分布模式进行比较，判断今天到底是不是一个普通的日子。如果今天确实是一个不普通的日子，那么在预

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

测中采用一个“比较保守的过滤器”，难道不是一种比较明智的做法吗？设计一个计量变化比率的方法（有一些标准的做法，从统计摘要，包括各种动差等，到综合衡量信息等方法，对概率的分布进行比较），并设计一个一般的动态更新方案，比采用60天移动历史数据周期的方法，更加灵活。

10.4 更多的黑匣子

在本章一开始，我们特意提到了摩根士丹利，这是因为它与我们的主题“统计套利”联系密切。但是除了摩根士丹利之外，还有很多公司从事交易数据的分析，并将好的交易方法整合到一些工具之中，向市场出售。2000年，高盛低价收购了在纽约证券交易所上市的 Spear Leeds & Kellog 专业交易商。高盛挖到了一座很有潜力的金矿，比摩根士丹利的数据库更具价值。2002 年，美国银行买下了 Vector 对冲基金的技术：“*……计算机算法会将特定的股票交易特点作为一个因素，如果美国银行自己的头寸在其中，就会产生买进与卖出的报价*”（2004 年 6 月的《机构投资者》，斜体字主要起强调的作用）。瑞士信贷第一波士顿银行（CSFB）聘请了声名远扬而且技术高超的对冲基金 D. E. Shaw 公司的一名员工，设计了一种工具“用来处理占了全部 40% 的交易订单”（2004 年 6 月的《机构投资者》）；雷曼兄弟和其他很多的公司也都参与到这一事业之中。

除了前面所谈到的一些发展之外，还有一个经纪商，米利都（Miletus），采用交易算法引入对冲基金进行交易，创造了一个规模达 10 亿美元的对冲基金。另外，交易技术带来市场的发展。2006 年下半年，高盛和至少两个对冲基金，将交易算法的方式引入到市场之中。由于广泛使用这些工具，可能有着一种新的系统性的交易模式加入到了这个市场之中。

10.5 市场紧缩

图 10-1 描述的是股票市场的买卖活动。在股票市场中，买卖双方都认

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

可某个价格，认为可接受这个价格进行所有权的交换。市场上存在很多的买方，也有很多的卖方。市场中存在很多令人兴奋的交易，同时都能达成一致并成交，也存在波动性。

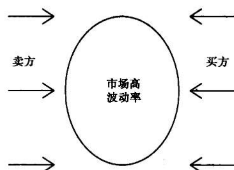


图 10-1 市场存在的方式

图 10-2 介绍了股票市场的买卖的模式。有很多单独的买方与卖方，他们聚集在一起，通过计算机算法来进行交易。这些算法在他们内部控制很多的交易订单，而在市场中剩余的交易则是受到了限制，并且交易量也不够活跃。在市场中，存在很多的买方和卖方，同样还都能达成一致并成交。但整个市场不如传统市场那么热烈，波动率也因此受到了一些限制。

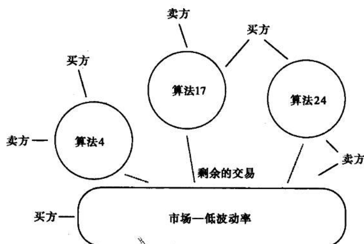


图 10-2 一个紧缩的市场模型

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net



第 11 章 统计套利的复兴

担心每一件事情会令人身心疲惫，这样做也不能达到预期目标，因为过度的担心，会对数据产生错误的认识，使我们对一个正确运转的系统，重复地进行维护，最终导致系统出错。

——博克斯

直到 2004 年年底，从事统计套利的人员，在这一年中都承受了很大的压力。投资者和评论员认为，投资绩效波动很大，但没获得市场想要的回报（在 2003 年）。投资者和评论员断言在当前市场状况下，统计套利的表现还会持续下降。投资者和评论员对于现实的复杂性（第 9 章和第 10 章对此有全面的解释）充耳不闻。

本章对这些负面消息进行了驳斥，并且论述了从 2006 年起，统计套利在绩效回报上的表现。第 2 章到第 8 章说明了传统统计套利机会和一些扩展的内容，包括一些模型化的方法和如何系统性地运用那些机会，以及由于市场动态变化的属性，给统计套利模型建立与管理的投资组合造成的巨大破坏等议题。第 9 章论述了对统计套利各种指责的理由，这些指责的逻辑是：市场的变化消除了部分统计套利的回报，变化是永恒的，因此统计套利的机会已经结束。这样的指责似乎是中肯的，但用来解释交易记录是不充分的。在更加复杂的现实中，市场有很多糟糕的表现，但经过深刻的反思之

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

后，这些表现不支持那些指责的理由。在复杂的市场中，存在着一些基本因素，没有摧毁统计套利这种交易模式。相反，在第 10 章中论述了更深远的市场结构的变化将创造出新的条件，让统计套利产生新的交易范例。涌现出新交易范例的驱动力、统计上的描述和可以运用的股价模式，都将在本章中进行论述。它是本书的一个结论，也为今后几年的发展拉开了序幕。

很少统计套利从业者能持续获得显著的回报。如同前几章所描述的那样，很多人并没有获得显著的回报。这个证据支持了下面两个主张：①传统的统计套利机会是不完善的；②新的机会有自身起源和发展。只有从自营交易的信息中才能揭示，支持上面两个主张的程度。通过分析股价的历史数据，所得到的证据强烈地暗示，高频率的交易方式（如一天之内的交易）能获得巨大的回报。在本书的讨论中，可以清楚地看到，开发利用那样的机会，需要不同于传统的平均回归模型。在本章中，稍后将描述一些这样的模型。

在一个交易日中，股价的变动模式没有出现反转的现象，而是具有某种趋势。在一个交易日之内，也会出现反转的模式，但这些模式看起来很难预测（尽管有人声称可以成功地预测）。对于投资组合来说，反转现象总是断断续续地出现，反转的先兆信号不容易识别。如果用“反转”来描述这种波动现象，是不恰当的，或许用“颠覆”一词更能说明这种动态的变化。两者之间的差别是很关键性的。反转过程先假定存在一个相对的平衡点，在一个干扰因素之后价格远离了平衡点，之后价格（或相对价格）又返回到平衡点（例如爆米花过程）。可以识别出平衡的力量。趋势过程和“颠覆”过程不存在平衡点这一假设，而是认为在一段时间内，先往某一个方向移动，然后出现一个往相反方向的移动。两者之间没有必然的联系（可以认为是一种无意识的转换过程）。向某个方向移动的持续时间和幅度，从描述与估计的角度来说是很关键的：向某个方向移动的时间足够长（可以认为有充分的时间，对于人们来说，每一个步骤都是可见的），而

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

且移动的幅度足够大，可以系统地利用这个机会，转折点的识别会存在一些时延。

模型成功的关键因素是了解市场中导致股票波动的趋势力量。报价的变动就是一个重要的因素，这个因素与价格摩擦无关，消除了在历史上价格来回移动（累积起来很大）的现象。根据第 10 章的描述，很多交易是通过电子交易所或者是经纪商的交易程序进行交易，专家设定价格的过程也就逐渐消失。现在，涌现越来越多“智能的”交易引擎，像 VWAP 或 TWAP 这样的交易方式占有了非常大的比重。很奇妙的是，一些传统的技术分析方法运用在每日交易趋势时，也还存在一些功效；不过，那些涵盖市场动力和算法交易策略的模型，能获得比较大的成功。

人们通常根据公司短期与长期的前景等主观看法，对公司的价值做出评估，这与根据平衡力量的方法进行评估大相径庭。新评估方法所采用的是以不带感情、没有利益冲突为基础的方法，持续地探查其他各种因素。这个过程是类似于机械论的方法，就像在一个地质学的过程中，水总是流向最低的地方。然而，这些由人类所定义的模型规则，不是宇宙中的物理定律，是会变化的。人类主导市场的行为，实际上也是所有交易的源头，因此噪声是无所不在的。尽管存在噪声，新的平衡力量不是为了寻找公平的相对价格，而是公平的（所有的市场参与者共同接受的）市场清算价格。这种新的范例可能是古老的经济学范例，即完全竞争的反转现象!!! 现在，在这样的思想之下，人们产生动态蛛网算法、博弈理论策略等想法，同时对于行为金融学研究的必要性进行重新定位。

各种算法不断地运用波动率。算法与算法之间的交易将取代过去在纽约证券交易所中面对面的交易，或者在电子化的市场中面对电子屏幕的交易。带有情感的交易方式，在很大程度上会逐渐消失，同样地波动率也会逐渐消失。尽管算法变成了新的焦点，我们还是不能忘记是人类驱动了系统。算法交易全都由快得令人难以置信的计算机进行管理，算法的设计除了被动的参与市场之外，对市场决策产生哪些主动的影响？由于算法的缺点、推翻幼稚

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

的想法所产生出来的机会、错误判断的误导等问题，需要持续地探索新的算法。在算法领域，这是一场战争。对于某些经理人和算法设计者来说，这项挑战性的工作是一种无法抗拒的诱惑。

当然，这仅仅是我的推测。

11.1 突变过程

从 2004 年开始，我们可以观察到价差的变化出现一种非对称的过程，这种过程的发散程度是缓慢的和连续的，但收敛（向原先的均值方向反转）得比较快，相比之下甚至有些突然。虽然它向局部平均值的方向移动，但并不是收敛于“均值”。这两个特征与爆米花过程相比较，发现爆米花过程是一种比较快速的发散和比较缓慢的回归过程。对于这两种不同的过程来说，第三个观察到的特征，即反转到相应平均值的程度，也是不同的：在新出现的过程中，往回移动的范围比爆米花过程更加不确定。

现在我们陷入了一个定义上的泥潭，所以必须小心地审查、详细地解释。

利用图 11-1，我们可以将经典的爆米花过程和这个新的过程进行比较。这个新的过程有几个显著的特性：从局部平衡点发散的行为比较缓慢而平滑；在大多数的情况下向平衡点反转的速度比较快一点；仅部分反转；快速而连续的移动过程，表现出与平衡点并不相同的局部趋势（在所有的原型示例中，平衡点是指一个常数。在实际中，以一个正向的长期趋势为例，我们可以沿逆时针方向将书页转几度，然后再去看这个原型）。

在这个新的“突变”模型中，最为关键的偏离程度是指一个期间出现了局部的偏离趋势，而这个期间在传统的状况中是处于常数状态。这个局部的偏离趋势，现在必须要进行描述，并且要加入到正式的分析中，因为它是驱动机会的一部分，也是成功地运用这种新的反转移动的关键工作。我们不能忽视它，仅把它作为爆米花过程中的噪声。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

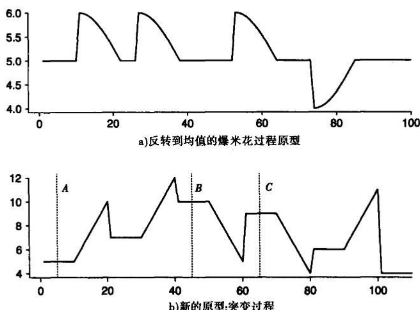


图 11-1 突变模型

“反转”的幅度的不确定，以及向相同方向多次的移动后往相反方向一个比较大的波动（一次事件，也可能是多次的小事件）等这些联合出现的现象，都是由于算法交易之间的交互作用所导致的（可能还有其他的驱动力量，但到目前为止还是很难以了解其他的驱动力量）。当价格反复地波动时，算法交易会很有耐心，对价格的波动产生缓和的效果——专业交易者会对十进制的报价变化很灵敏，毫无疑问这些问题会在一些程序的算法中有所反映。当报价单位为 1/8 美元时，报价单位比较大，价格波动难免会有一些惰性，现在是报价单位为 1 美分，会反复出现价格波动。当人类交易者主宰订单流量时，一开始将报价单位缩小到一美分会产生一些荒谬的可获利机会。但算法交易的耐心和纪律渐渐取代了交易者认为的干预，因而改变了交互作用。现在似乎比较清楚了，这样造成的结果就是所谓的“突变”过程。

请注意，如果将爆米花过程模型应用到突变过程，相对价格的演变会呈现出零回报的结果。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

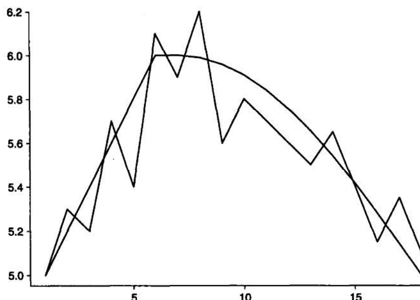


图 11-2 用突变波动来扩展爆米花移动的细节

看到这里，我们自然而然会产生一个疑问：当面对比图 11-1 中从 *A* 到 *C* 更长的时间，会发生什么呢？刚刚所给出的描述是：向某个方向一连串的移动，不时被相反方向的移动所间断，同时在相反方向的移动也呈现相同的情况。图 11-2 就是这样的，并且图 11-3 也是类似的。这些图形有点像用放大镜看爆米花过程。这些图形像蒙上了一层阴影似的，没有什么意义，没有什么经济价值。比较适当的解释是：由于时间长度的扩展，因此使得爆米花过程中的微小波动，在时间维度上变成跟原来的爆米花移动一样。因此在整个爆米花移动的过程可能包含有六七个（甚至更多）突变过程。在一个动态的市场中，这是一段很长的时间。在理想的条件之下，爆米花的回报还是降低了好几成。但是如果采用像图 11-1 那样真实的波动图形，其回报率可能更加差。以刚才建议的时间段来说，局部均值变化的大小超过了假设所能允许的程度，使原先的爆米花过程失效。我们能做的就是如图 11-4 所示，用一个比较平滑的曲线来表示爆米花过程的平衡点，但偏离平衡点的程度降低了我们的兴趣，把一个原本可利用的结构，转换成了一个只有教科书或期刊才感兴趣的题目。如果考虑较长的持续期间，交易由市场的基本面来决定。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

由于统计爆米花过程不是根据基本面的分析所做出来的结论，因此预测的准确性和回报都将随之消失。

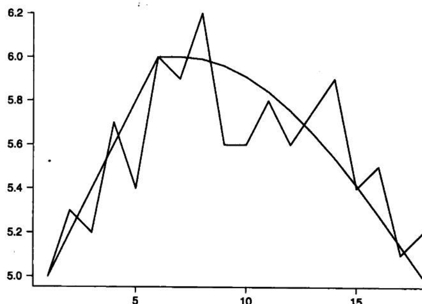


图 11-3 将爆米花过程进行扩展的一个异体

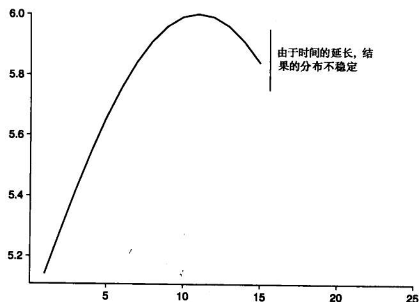


图 11-4 扩展的爆米花过程会有不确定的结果

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

11.2 突变预测

与针对爆米花过程的预测结果进行对比，突变过程无法准确地预测反转量的大小。将爆米花过程，应用到爆米花数据中，而突变的过程，则应用到突变数据中，根据变动的原因进行分类。那么捷径在哪？例如在 2002 年上半年，其图形像是一个爆米花过程，2004 年下半年则属于突变过程，而中间的 18 个月，剧烈的变化造成破坏性的影响，因此对其中很大一部分交易来说，两个过程所得到的统计均值为 R_2，交易结果是类似的。对于交易来说，这样观察的意义就是，假如在合理的一段期间中，两种过程所产生的交易数量是类似的，交易的总变动也是类似的，那么回报率的期望值也应该是类似的。当然，在实际中不能如此简单地处理。当突变过程对价差运动的特性进行描述时，比爆米花过程更准确一些，那么价差波动率的幅度也会显得比较低（见第 9 章）。在 2003 年之前，当时的平均波动率差不多是 2004 年下半年的两倍，爆米花过程对于价差行为的描述是比较准确。在 2004 年下半年，当时的平均波动率大约是 2003 年的一半，突变过程是一个更准确的模型。这些并不是巧合。这些都将对市场的理解加入到算法交易之后导致的结果（见第 10 章中的描述）。

总体方差变小导致模型预测也变小——表面上看来，这体现了回报减少与方差收缩两者之间的关系。但当移动持续的时间比较短，或者是比较高频率的移动，面对的情形就改变了。在这种情况下，虽然个别交易的收益降低，增加交易的数量，回报也会增加。交易数量增加会造成的交易成本增加。向经纪商和使用交易算法所缴纳的费用，会逐渐下降，这是一个必然的结果。

在这里必须要回答的一个关键问题是，如何系统性地利用突变信号来进行交易，才能产生值得的、具有经济价值的回报？

在理想的状况下，人们想要在突变过程中出现突然的变化之前，识别它的开始点，并且在没有对市场产生冲击的情况下有充足的时间进行交易，并

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

且很快就能识别出突变过程的结束，以便取得最大的利润。到目前为止，还没有一个很简单的识别方法，但是基于持续时间计量的近似方法已经建立起来了。

如图 11-5 所示，显示了突变原型的成长与下降（或者下倾和跳跃，如果你更喜欢反转）情形。将注意力放在突变移动的建立过程，我们就可以识别出一个跟持续时间有关的规则，这个规则可以在趋势产生之后的第 k 个周期，找出交易进场的信号。只要经历几个周期，可以知道趋势已经开始了。在“突变”转向之前，通过统计分析，可以得到趋势持续时间的分布。因此可以将交易的进场信号，设定为这个分布的某个固定值。例如将固定值设为 80%，就可以得到一个不错的操作规则。

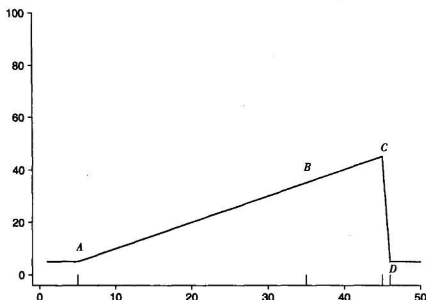


图 11-5　突变移动原型

如果能及时地识别从发散到突然反转之间的变化和断点，那是成功利用突变过程的关键因素。与爆米花过程进行比较，在突变过程中，交易进场的时机在统计上来说更严格。突变反转的速度相对较快，太晚识别出突变交易时机的损失，会远大于爆米花过程。在图 11-5 中的 C 点是峰值的最高点，在 C 点之前未能进场意味着错过了整个的机会。在图 11-3 中，爆米花过程

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

与之有很大的不同，太晚进场会降低交易的回报，但影响有限而已。要对突变过程进行模型化交易，必须要有更高的警觉性。

11.3 趋势变化的识别

在统计套利的发展中，有很多的文献讨论变化点的识别，有很多有趣的模型和许多令人着迷的方法。在这里我们的目的只是进行一些简单的比较，虽然如此，这仍然是一个具有挑战性的工作（如果对金融数据的模式识别不具有挑战性，那么我们也就不用辛苦地撰写和阅读文献）。在统计过程的控制领域中，Cuscore 统计量是一个极为有效的方法（博克斯与卢西诺，1987）。

首先将突变过程添加到一个上升的数据序列之上。图 11-6 显示了一个基本的趋势，其斜率为 1.0，同时一个斜率为 1.3 的突变移动，从时间 10 的地方开始出现。让我们来看看，在这个很容易理解的示例中，趋势变化的 Cuscore 统计量表现如何？为了识别出趋势中的变化，将 Cuscore 统计量定义为：

$$Q = \sum (y_t - \beta t)t$$

其中 y_t 是一系列的观测值，β 是斜率（也就是观察的时间序列值在每个单位时间中的变化率[⊖]），而 t 是一个时间指数。Cuscore 值形成的图形如图 11-6 所示。尽管我们看过许多时间序列图形，当我们看到这个看起来不可思


注 释

⊖ 在模型中，采用参数的变化值，比起单纯采用时间索引序列值，具有更广泛的应用价值。但在这里，我们把焦点放在时间索引序列值上。在空间模型中，观察到的序列值是用地理空间上的位置作为索引，而不是按照时间的顺序作为索引。这种空间的模型应用在很多场合，从地质学到地震学（经常同时采用时间与空间作为索引）到生物学（脑电波图根据头部不同位置，以及每个位置随着时间变化的发展，找出其特殊的模式）等科学领域。在股价分析中，某些交易算法（参见第 10 章）和某些统计套利者会采用交易成交量作为索引。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

议的方法时，还是很吃惊。这种统计识别方法揭露和展示了斜率上的变化所呈现出来的证据。当斜率从初始值 1.0 增加 30% 变为 1.3，其变化的幅度应该是很明显的。30% 已经很接近 1/3 的程度，是一个很大的变化，应该引起我们的注意。但这个图形让我们有一个非常不同的感觉。如果我们在图 11-6a 中画上一条延长的虚线，我们很难在时间 10 的地方注意到已经出现了变化。这个图形从视觉上看起来是很协调的。在这样的图形中如果配有文字进行说明，那些文字可能更容易受到大家的注意。

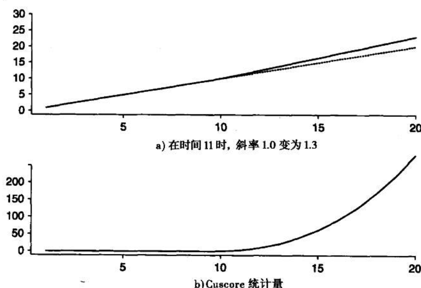


图 11-6 趋势变化的识别

Cuscore 统计量将常数转化为指数函数，这样戏剧性的转变，有效地揭示了状态的改变。当观测值并没有落在指定的数学曲线上时，Cuscore 统计量的表现如何呢？图 11-7 加入了随机的噪声（自由度为 5 的学生 t 分布，其尾部的分布会比正态分布厚一点），将其放到图 11-6 中所描述的序列之中。如果说之前斜率增加这样的变化，在视觉上已经很难识别，那么现在在视觉上更是不能识别。Cuscore 统计量又表现如何呢？

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

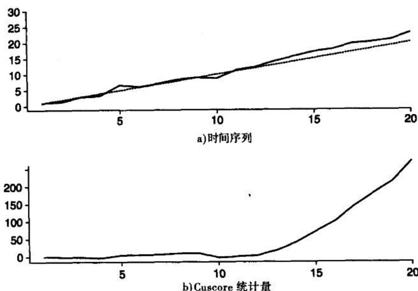


图 11-7 对于带有噪声的数据来说，Cuscore 统计量识别趋势变化的情况

在图 11-7 中，Cuscore 统计量的图形比之前的更加戏剧化。对此，我们更是兴奋。通常很少用眼睛进行分析，因为很容易产生疲劳。在时间 15 之前，Cuscore 统计量出现了趋势增强的信号，而且在一或两个周期之后，也确实发生了。

11.3.1 利用 Cuscore 统计量识别突变过程

在最前面的例子中，趋势采用的是常数，Cuscore 统计量的系数 β 为 1.0。不幸的是，我们都很难量化金融序列的变化率，必须使用原始的观测值。如果想在一个价差序列中，识别出其中的突变移动，需在突变过程发生之前，得到一个相应趋势值。我们最先想到的可能是，利用第 3 章中所建议的指数加权移动平均（EWMA），来计算局部的平均值。但这个想法显然存在“先有鸡，还是先有蛋”这一问题。不管是采用 EWMA 还是其他的公式计算出的局部平均值，在突变过程一开始出现时，掺杂到局部平均值的计算之中。如果计算斜率时，所采用的数据只包含突变过程发生之后的数据，而没有包括突变过程之前的数据，Cuscore 统计量不可能识别出斜率的变化。我们所需要的是，在变化发生之前得到一个趋势的估计值，如果出现变化的

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

情况，这个估计值也能产生相应的变化。由于变化发生的时间点是未知的，在这种情况下，我们能做些什么？

下面介绍的两个简单策略对此会有一些作用。第一个策略是，在突变过程确实存在的前提下，估计当前的趋势，我们可以利用一段比较长的时间周期中（例如将一般突变过程所持续的时间乘上几倍）的价差交易数据，观察其突变过程所呈现出来的经验分布[⊖]。第二个策略是，采用一个斜率作为当前趋势的估计值，这个估计值可以用 EWMA 来计算。对于研究的序列来说，这个估计值是一个比较合理的值。修改后的 Cuscore 统计量公式可以表示为：

$$Q = \sum (y_t - \hat{\beta}_t t)t$$

其中的 $\hat{\beta}_t$ 是当前斜率的估计值。在附录 11A 中，我们给出了一些关于 $\hat{\beta}_t$ 的来历，对于识别股价趋势变化的 Cuscore 统计量进行仔细地审查。

操作上来说，Cuscore 统计量运作得很好，但如果还想要更早地识别出突变过程，应该还有其他更好的识别程序。其中一种可能的方法（在附录 11A 中有介绍）是采用一个延时局部趋势估计值，可以避免“先有鸡，还是先有蛋”这样的问题。因为我们“已经知道”突变过程开始之后，可以在 5 个单位时间之内，把它识别出来，所以可以在估计相应的趋势时，先去掉至少 5 个最近的序列观测值。这样做或许是合理的。

为什么不取 5 个以上的观测值，用于延迟计算 EWMA 值，以达到“安全”的状态？（就技术上而言，当突变过程以“温和的”方式出现，即便在有噪声的情况下，也可以提交识别突变移动的概率）。这个问题与模型设计


注 释

⊖ 从过去的数据中识别出突变过程，比实时的识别要容易了许多。可以利用历史数据识别出某些移动过程，然后将其指定到一个类别之中，作为研究和描述规则时作为标本来使用。在确认该过程属于哪种突变过程前，通常就必须实时做出决策，并进行交易。在获得利润，或者产生损失之后，我们总能确认突变过程的种类。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

者的设计艺术和识别者的执行效率是密切相关的。当你从事于这项研究时，需要考虑如下几项重要的事情：

- 一般的趋势与发生突变过程之前的趋势之间的差异，呈现什么样的分布？
- 突变过程反转的幅度，呈现什么样的分布？
- 突变过程趋势累积的持续时间和突变反转的大小之间有什么样的关系？一般的趋势与发生突变过程之前的趋势之间差异的大小，和突变反转的大小之间有什么样的关系？
- 想要捕获具有经济价值的突变过程，应该具备有怎么样的条件？
- 如果错误地识别突变过程，会造成多大的损失？

祝大家能顺利捕获突变过程！

11.3.2 突变结束了吗？

如果价差序列返回到（局部的）均值，只要超过均值一点点（考虑到随机共振的作用），爆米花过程就算是结束了。什么时候突变过程才算结束呢？这个问题目前最稳当的答案是：如果在突变过程发展的相反方向识别到峰值，那么出现峰值的时点往后取一段固定的时间作为突变过程的结束。如果突变过程的发展是增强原来的趋势（如同之前的示例），那么这个突变过程结束的时点，就应该是突然下降的时点。

在前面的统计套利模型化过程中，我们并没有一个成功的方法来解决这个问题。如果突变过程仅仅是一次性的大型波动，或者是跨越好几段期间的趋势变化，应用 Cuscore 统计量进行监视，那么只要它一出现就能识别。如果采用前面第二个修改过的 Cuscore 统计量，来识别移动的特性，我们所监视的区间，是从估计的突变过程起始点开始，到最后一个观察值之前的一或两个周期为止。通过 Cuscore 统计量，我们可以寻找出突变过程一开始的发展，但如果遇到相反方向的突然的变化，则会延迟一段时间才能辨识出来。如果我们将最后面的那些观测值包含进来，就会出现前面所描述过的状况，Cuscore 统计量无法识别出突变过程的起点。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

因此到目前为止，如果想要为交易制定一个退出的规则，最好方式就是考虑突变过程的持续时间与突变移动的大小，然后将各方面的情况结合起来，决定突变的结束点。这种做法有个很重大的危险是如果等待的时间太久，又出现了另一个突变过程，就会损失第一个突变过程所获得的利润。在模型与交易规则的调整方面，还有很多可改善的空间。

11.4 突变过程在理论上的解释

算法只是一些基础的公式，不同于人类的意识。大多数的交易者的行为与他们的模型并不一致。但是算法会默默地保持前后的一致性，不会有什么想象力。市场中常常会出现一些随机出现的行为，不管是单纯根据个别算法进行分析，或是考虑算法的交互作用，这些随机行为都是无法预测的。

在图 11-8 中，研究一下突变过程的表面现象。

[注意] 我们之所以选择“突变过程”来命名这个新的反转模式，而不把它叫做爆米花第二，或者是其他的名称，是因为这个名称比较容易被人记住，而且它确实比爆米花过程这个名称，更能生动的描述动态的移动过程。对投资者行为的模型进行解释，可以描述这种新形态发生的原因。这些模型的发展和移动的描述与开发，必须分开来看。算法与算法之间相互影响的原理，以及交易算法越来越流行，直接取代了人类的交易行为。这些现象是毋庸置疑的。这些都是事实。股价的历史数据也是不可争议的事实。我们在那些历史数据中识别出来的模式，都是具有争议性的：在我们所能拿出来的示例中，都存在很多的噪声。

无可否认，利用爆米花过程与突变过程表现出来的动态过程，建立的交易模型都采用追踪历史数据的方法。那些历史数据证明了模型的功效，并支持了模式描述的有效性，但是它却没有办法证明，为什么模式会以这样的一个方式存在。人们发现爆米花过程已经有很长的一段时间，它被广泛地利用

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

在多种情况下，并被作为一种合理的理论，只是很多人并没有注意到。当旧的模式失效（失去了经济上的开发价值）时，新出现的模式不必是合理的。只要统计套利者开始发现了新模式，而且这样的状况一直在持续，并且可以得到一定的回报，投资者就会从怀疑的态度（对损失的恐惧）转移到充满希望的态度（贪婪）。

尽管突变过程中出现的反转现象的原因不需要一个合理的解释，但能了解一下驱动这种动态变化的市场力量，以及那些力量如何交互作用，如何产生出随机出现的模式，得出一个令人信服的理论，是很有必要的。在文中所呈现的简单的突变理论模型，作为识别市场力量的一种可能的方式，其影响力在日益增长，取代了一些旧的行为与交互作用。这种现象也是可以理解的。对于目前已知的情况来说，突变模型似乎是一个好像有点道理，但它并不是一个可以用来进行预测的正式模型。阿诺德（VI Arnaold）在其《突变理论》一书中刻薄地评论道："关于突变理论的文章，很多只是因为它与过去的看法不同，而目前所能看到的相关出版物只是比较新鲜而已。"你应该要提高警惕。

阿诺德更进一步地评论说："大多数严肃的应用中……在突变理论出现之前，结果就都已经知道了。"本文想要强烈说明的是，尽管阿诺德在 20 年前写下了这些话，这个模型所呈现的表示方式与解释即使还是正确的，模型设计者仍然有机会可以运用一些突变理论之外的方法，构建出更好的（更精确的，更有说服力的）模型。事实上，我本身正在从事的研究，就运用了博弈理论上的工具，对交易算法的交互作用进行模型化。这些相关的工作，目前都还只是处于很初期的发展阶段，所以并不适合进行说明。最后，再次引用阿诺德的话："关于股票市场交易者行为理论，就跟原本的假设前提条件一样，其结论只是具有启发性的意义而已。"我们所根据的前提假设不仅仅只是具有启发性的意义，算法对算法的相互作用现象，以及算法愈来愈增加的支配能力，还有人类直接参与的作用越来越少。任何人都可以从股价历史数据中，发现这些特定的模式。虽然如此，但是对于市场交易者来说，突变

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

模型确实只是一个具有启发性的模型。按照阿诺德对启发性的说法，或许我应该将我的模型叫做蝌蚪定理（Tadpole Theorem），有个明确的意图就是，这个名字里有个 Pole（作者的名字）！

突变过程是先经过一个缓慢积累的趋势之后，接着又突然发生逆转的过程，可以通过一个二维空间的连续移动来描述。在突变过程这个理论中，其中的一个维度对应“正规”因素，而另一个维度对应“分裂”因素。当分裂因素处于很低的水平时，正规因素中的变动会在图形表面形成平滑的变动。当分裂因素的水平很高，正规因素的移动所形成的图形，会落在两个分开的区域中，会出现一个不连续的状况（突变跳跃）。这种不连续的状况是不对称的：当分裂因素值处于一个常数水平的情况下，“往上”与“往下”的现象分别是不同的正规因素数值。这就是滞后现象，一般解释为惯性或者是阻力（图 11-9 显示了突变现象的一个横截面。它是在分裂因素处于一个高水平的常数值情况下，以平行于正规轴的方向，呈现的横截面图形，展示了非对称的跳跃过程）。

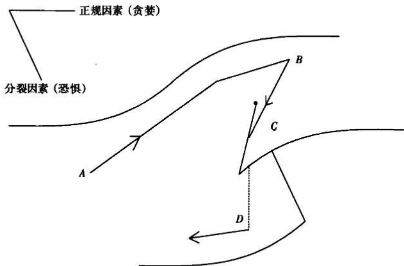


图 11-8　突变过程的表面现象

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

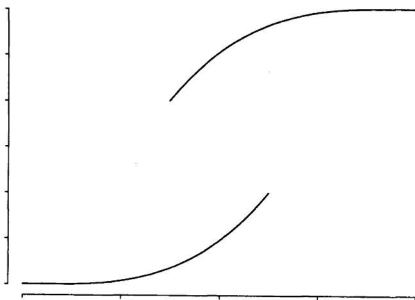


图 11-9 突变过程在高水平的分离因素下，所得到的横截面

这是对二维的尖端突变模型（Cusp Catastroohe Model）的经典描述。应用到股价发展时，可以用正规因素值来识别“贪婪”的程度，而分裂因素值则可以用来识别“恐惧”的程度。我们看一下这个移动过程，它从 *A* 点开始，伴随一个水平相当低的“恐惧”因素。随着贪婪程度的逐渐增加，价格向 *B* 的方向渐渐平滑地发展。当价格更进一步的移动到 *C* 的位置（价格在图形上沿着分裂因素轴，往恐惧增加的方向移动）时，恐惧就开始传染到每个参与者的身上。最终，恐惧超越了贪婪，成为具有支配性地位的主要因素，而价格也就出现了一个不连续的、快速的回升。恐惧的特性是什么呢？价格从局部趋势中，突然出现不连续性的、发散性的变化，并不是根本的变化，而是由于过度的利用买方或卖方而造成的变化（实际上，算法并不会感到恐惧，对此也没有任何的经验，不会依照情感行事。如果用非正式的方式来描述的话：这方面是一个在开展中的工作。我不打算用任何的方式，建立相关的理论。正如你所见，在观察价格的动态变化方面，我也还在寻找着适当的说明与解释）。再重复一次：算法对经验而言，是无意识的。不过，

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

算法确实会在价格移动的动态变化中所学到的知识（参见第 10 章）。所有的这些知识，以及目前市场波动的信息，会被送进一个计算公式中，从而呈现出“恐惧—拉回”的现象。

通过对恐惧与贪婪因素的描述，将参与者（有买家、卖家及专业交易者）的交易算法表现出来。贪婪轴衡量的是，影响着交易者与专业交易者的最大贪婪程度。无论谁带有最贪婪的看法，就会对价格移动与交互作用产生支配性的影响。以类似的方式，恐惧轴也可以衡量出，传染到参与者最大的恐惧程度。

当专业交易者看到市场中的购买压力时，就会开始提高卖方报价。交易算法为了顺利完成交易，一般会允许一定的价格上浮空间，让专业交易者把价格抬高。这样一来，专业交易者的贪婪会逐步增加，一步一步将价格抬高（可能会加快价格变化的速度，不过在这里，不需要描述具体的细节）。随着这些相互作用持续，价格会涨得更高，一直到交易算法认为是到了该停止买进的时候。算法会利用历史数据进行校验，“预测”需要多少钱才能完成交易，然后算法就会以不带任何感情的方式，凭着耐心执行交易。一旦购买压力消失，专业交易者的贪婪立刻就会转变为恐惧。如果买方没有买入动作，卖方维持的高价只会造成在市场上没有成交的机会。卖方将扩散恐惧。于是价格就会突然下跌（跟价格逐渐上升相比），期望点燃买方的兴趣。

有人可能会问，为什么不是逐渐地下跌呢？这是因为人们对于恐惧的反应，跟对贪婪的反应（怕卖得太便宜，或者是怕买得太贵）是不相同的。因此算法也会有不同的反应。请注意算法和程序代码都是由人所设计出来的。买方会比较有耐性，等待一个明显的下跌。因此，在很多的情况下，没有什么介于中间的交易，价格只会加速下跌，就像我们观察到的灾难性下跌。

如果买方的耐心耗尽，就会重新开始一次新的循环。新的起始价格可能会高于上一次的循环的起始价格，这是因为短期的突变现象只会反映出部分的效应。狂热与贪婪又迅速地出现，在足以支撑价格上涨结束之前，价格竞

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

赛会继续。在现实中，如果对于恐惧只是暂时的，那么价格将很快恢复到平衡点。

算法相互作用，以及其所导致的股价行为，会对价差产生什么样的影响呢？影响是很直接的。股价会分别以它们自身不同的规律进行波动。个别股票的突变波动，很自然地产生价差之间的突变波动。价差的动态变化方式是不同的，波动的幅度也是不一样的。但是基本的特征是相同的。

11.5 风险管理的含义

如果想要成功的管理很多个统计套利模型，一个很有价值的风险管理工具就是所谓的最低可接受的回报率（hurdle rate of return）。每个模型的预测函数，都可以为任何打算的交易，提供出一个明确的回报率期望值。管理者一般会指定一个回报率的最小值，在交易之前，必须先确定回报的期望值高于这个最小值，避免进行了一堆交易，聚集起来却是输家。在感知到一般风险增加时，例如波动率增加的话（不管是实际波动率，或是预期的波动率），在实践中，标准的做法是提高最低可接受的回报率。设计这个预防的动作是为了要在分散的趋势还很强大的时候及时退出，避免反转交易过早进场，避免损失初始投资，因而增加回报。如果我们考虑的是一般性的波动增加，而非聚焦在特定的市场区段或股票上，那么这种战术是恰当的做法。这种战术是对不同模型进行一个概括性的描述（当然，如果有理由需要进行这样的考虑，这个战术也可以直接用在特定的金融市场，或者是其他股票集合上）。

对于爆米花过程来说，基本的预测函数是一个常数，这个值可以随时用 EWMA 的方法来计算出来（资深的模型设计者会采用局部的趋势成分法，按照移动所需的时间范围进行调整）。当价差突然出现，以用价差与预测值之间偏离度的比例，计算出回报期望值。如果预测波动率将要增加时，我们就会预测价差偏离度也会增加。这样的话，等待出现比较大的价差偏离度，显然是一种明智的做法（波动率缓慢增加，而不是突然增加，那么模型自动

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

地管理波动率，将其反馈到模型的动态重新校验程序之中。在这里，我们所关心的场景是在一个较短的时间里，偏离度增加到超出模型自动调整能力的范围之外。这种情况是一种具有风险的场景，而不是一般的动态的发展过程)。对于一个以“数据分析”为基础的模型来说，这种以“预测”为基础的数据是不容易理解的。这需要模型设计者进行一些沟通说明。

如果考虑的是突变过程时，波动率突然出现非正常地增加，需要考虑哪些风险？第一感觉应该是与爆米花过程一样。突变过程就是在一个缓慢的分离动作之后，突然发生收敛的情况。所以，在波动率突然变大的状况下，同样有需要重新调整偏离度的幅度，这与爆米花（或其他反转）模型是类似的。但进一步地思考，发现第一感觉是存在问题的。自从 2004 年上半年开始，突变过程是对于局部价格与价差移动一个比较好的描述方式，市场（与价差）波动率水平也出现了历史的低点（见第 9 章）。当波动率突然变大的时候，会发生什么事，我们没有任何经验。重新调整局部突变过程的幅度可能是一个结果。但它可能是其他完全不同的结果。可以得出一个结论：增加波动率会使与突变过程相关的方法失效，在一定程度上，使得识别与利用突变过程的能力丧失，导致它回到了爆米花过程。如果算法之间的相互作用确实会导致价格产生动态变化，这一观点是正确的，那么这样的一个发展是否在理论上产生突破？是什么导致波动率突然变大？当然是人。算法只是工具，最终，是人驱动了这个过程。在这点上，我们对于投机领域的作用是很大的。这里有几个提示点，可以引导你进一步思考：

- 在一个局部的趋势中，等待更长一点时间，利用持续时间这一标准，而不是回报率期望值这一标准（能不能与回报预测一起考虑?）。
- 如果为一个比较大的趋势过程，等待更长一点时间，那么机会将会变得很少。在突变过程中，不是针对平均值做出反应，而是针对一个旧的、实践中甚至是不相关的基准的水平值做出反应，因此对于突变过程的反应方式并没有改变。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

11.6 结束

在实践中，当价差波动率增加时，旧的反转回归到爆米花的过程，与新的突变过程，形成一个连续的混合过程。对这种新过程的研究仍处于初级阶段。

传统价差波动率驱动着反转现象，其作为一个系统性的回报来源，可能会复兴。利率的提升，创业者承担的风险逐渐增加，或是突然的衰退导致市场混乱，都是潜在的驱动力。这种潜在的驱动力总是存在的，但是机会是有限的。由于报价单位采用十进制所产生的冲击力，交易的方式（VWAP，或是其他的算法）更有耐性，还有其他竞争因素（第 9 章）的影响，都会限制回报率。

新的示例还是有一定的前提。但是，它还是不能令人兴奋地尖叫——或许这种令人兴奋地尖叫，就像圣诞老公公雪橇上的铃铛，只有相信的人才能听到吗?

附录 11A 理解 Cuscore 统计量

为了识别出趋势的变化，统计学家开发出 Cuscore 统计量方法，应用在工业过程控制中，其目标是只要某个流程需要调整时，尽快地发出预警。以生产某个给定直径的滚珠轴承作为示例进行讲解。目标的均值（直径）是已知的。抽取滚珠轴承的生产样本则是依照生产的顺序进行，并且进行测量，然后计算出样本的平均直径，并将抽样结果画在一张图表上。由于生产产品的机器会发生磨损，每隔一段时间，总会发生偏离目标均值的情况。滚珠的直径开始增加。Cuscore 统计量的图形非常迅速地揭示生产机器开始出现磨损（在实践中，样本的直径范围也会被监视，不同类型的机器磨损会造成不同类型的后果）。

在工程领域中，就像是滚珠轴承那个例子，目标值是已知的。因此，识

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

别趋势变化的 Cuscore 统计量，$\sum(y_t-\beta t)t$（斜率系数为 β），就是一个给定的值。正如我们在第 11.3.1 节提到的，股价波动的情况是不同的。我们不能预先设定一个目标价格，从而建立一个因果机制（尽管进行了预测分析）。因此，为了识别出价格趋势的变化，需要对趋势变化进行估计。计算最新的趋势估计值可以采用局部趋势估计方法。监控局部趋势估计值的一个时间序列，可以发现趋势变化的直接证据。

在附录中，我们将检验用于识别趋势变化的 Cuscore 统计量的一些技术细节。这里的研究将揭示 Cuscore 统计量是如何运作的，以及使用局部估计值存在的一些问题。后面的这个问题特别重要。深入了解趋势的估计值对 Cuscore 统计量产生什么影响，对于成功应用这种识别方法，及时的利用突变过程，是一个非常关键的问题。如果不能及时进行鉴别，在现实中，就不会捕获有经济价值的机会。

在图 11-10 中，ABC 这条线是一个趋势变化的原型。第一个线段 AB 的斜率为 0.5，而第二个线段 BC 的斜率为 1.5。虚线 BD 是直线 AB 的延长线。虚线 AE 与直线 BC 平行，斜率也是 1.5。我们会使用这些直线（假设它们是无噪声的价格轨迹，希望有所帮助），来说明在不同的趋势假设条件下，会对 Cuscore 统计量产生什么影响。在这个无噪声的理论模型中推导出的结论和知识，会为我们研究有噪声的价格序列时，指引方向。

Cuscore 统计量 $\sum(y_t-\beta t)t$，就是将观察序列 y_t 的累积偏差相加，以及假设斜率为 β 的期望值。在图 11-10 中，将 Cuscore 统计量转换成直线 AD 在 y 轴方向上垂直偏差的和。第一个观察到的知识是，所有在 AD 上的点对于 Q 的贡献都是零。如果斜率没有发生变化，Q 近似为零。

当斜率发生变化，观测值就会偏离基础模型（也就是斜率没有变化）的期望值。顺着直线 BC，y 的值超过了直线 BD 的期望值，随着时间不断地增加。在图 11-11 中，我们根据 Q 值的累计偏差，就可以得到如标签 1 所示的图形。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

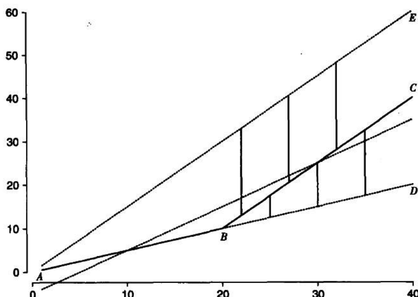


图 11-10 趋势变化的原型和 Cuscore 统计量相关的细节

现在，假设我们事先不知道直线 *AB* 与 *BC* 的斜率，也不知道在 *B* 点斜率发生了变化。假设我们最好的理解是，从 *A* 点开始应该出现一个 1.0 的斜率，如直线 *AC* 所示（在图中并没有画出来，以减少混乱程度）。在图 11-11 中，Cuscore 统计量显示为标签 2 的图形。Cuscore 统计量的图形在视觉上再一次令人吃惊。根据这个假设的基础模型，得到的偏差序列，可以明显看出趋势的斜率发生了变化。第二个示例揭示变化已经发生（在 Cuscore 统计量图形中变形的点），也揭示其他信息，例如数据序列一开始的斜率比假设值小一点，然后转换为一个比假设值大的斜率。

到目前为止，你可能对接下来的几个步骤有一些（或者更多）模糊的概念。回顾一下图 11-2，前三个突变过程组合形成了一个强烈的正向趋势；接下来的移动则组合成为一个不确定的衰退趋势。在实践中，我们如何从长期的趋势变化中，根据 Cuscore 统计量提供的一个合理的契机，识别潜在突变过程？本文给出了一个答案，就是使用一个局部的趋势估计值。当用一个局部的估计值来取代已知的常数趋势，让我们来检查一下 Cuscore 统计量的表现。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

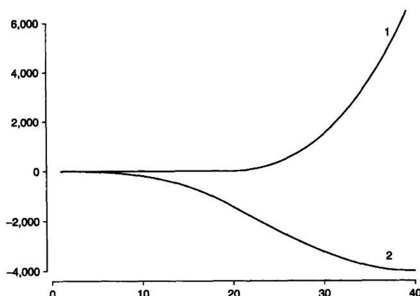
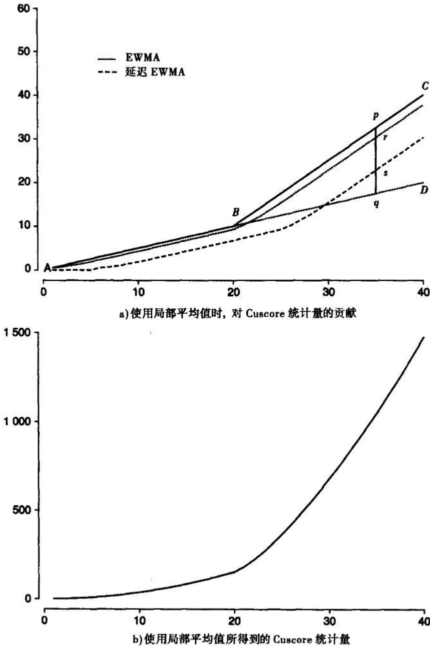


图 11-11 $\beta=0$ 与 $\beta=1$ 的 Cuscore 统计量

在图 11-12 中，EWMA 一直低估了实际的序列值；这是移动平均值或者加权平均值的显著特征，它们的目的不是用来描述持续的趋势。因此从序列的一开始，Cuscore 统计量“一直努力跟上”形势的发展，所以出现渐渐变大的情况。斜率的变化被抓取出来，Cuscore 统计量增加的速率也被抓取出来，同一个已知常数趋势的 Cuscore 统计量进行比较，它们的变动就比较缓慢。其直接原因是由于在斜率变化之后，应用 EWMA 方法。在图 11-12 中，Cuscore 统计量就是新的斜率与旧的斜率（直线 $BC-BD$ 之间的垂直差异）之间的差异，如图中 $p-q$ 所示。以估计的水平来说，从初始趋势线 AB 延长到 BD 产生的 $p-q$，会被 EWMA 产生的 $p-r$ 取代，这个偏差值更小。这是类似“先有鸡，还是先有蛋”一类的问题。我们需要在那个未知的变化点刚刚出现时，快速地识别出那个变化点！

斜率变化之后，更新局部平均的估计值，会降低 Cuscore 统计量识别变化的灵敏度。也就是说，延迟局部平均值的更新，可以恢复 Cuscore 统计量的灵敏度。如果在 Cuscore 统计量中使用延迟的局部平均估计值，会发生什么呢？回到图 11-12，在 Cuscore 统计量从 $p-r$ 增加为 $p-s$，非常接近我们

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net



a) 使用局部平均值时，对 Cuscore 统计量的贡献

b) 使用局部平均值所得到的 Cuscore 统计量

图 11-12

想要的 $p-q$。不幸的是，这个变化没有消除“先有鸡，还是先有蛋”的问

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

题，只是重新布置了鸡舍！这个方法对于 Cuscore 统计量的贡献确实比较大，同时对变化之前的贡献也很大。因此，准确地区分 Cuscore 统计量图形中的变化，是不能轻易实现的：这将导致过早地发出信号。降低延迟的程度（我们使用 5 个时间区间，是因为在价格的历史数据中进行多次关于突变过程识别分析之后，认为在 5 个时间区域内，可以识别出具有经济价值的突变过程）可能会有所帮助，但是只要从无噪声的原型转变为带有噪声的真实数据，就会变成没有什么希望。

我们寻找的是数据中的某个东西，它能在趋势变化出现之后，迅速并且稳定地识别出一个真实存在的变化信号。在图 11-12 中，EWMA 的图形能迅速地对趋势变化做出反应。按照 EWMA 方法得到的局部趋势估计值，可能就是一种灵敏的诊断方法？图 11-13 显示了一个斜率系数估计值的图形，这个斜率系数的估计值是在最近的四个时间区域中，采用 EWMA 方法计算出平均的变化值：

$$\hat{\beta}_t = 0.25(EWMA_t - EWMA_{t-4})$$

这个估计值显示了一个以 EWMA 为基础的 Cuscore 统计量。不幸的是，只要在原来的序列中加入少量的噪声，斜率系数的估计值就会持续恶化，不能作为一种灵敏的诊断方法。当趋势是常数的情况（如图 11-14 所示），这种方法的灵敏度会好一些。

在这里，我们有两个识别趋势变化的方法，一个是运用局部平均估计值（EWMA）计算 Cuscore 统计量，另一个是基于 EWMA 计算局部斜率系数估计值。这两个方法看起来都有前途。Cuscore 统计量可以识别出变化，但有些缓慢。斜率估计值也可以识别出变化，但数据是带有噪声的，也会有些缓慢。把这两者结合起来会提高效率吗？在我们讨论之前，先让我们来回顾一下到目前为止涉及的 Cuscore 统计量。图 11-15 展示附录中介绍过的所有 Cuscore 统计量，应用到无噪声的趋势变化序列中的情形。Q_{theo} 是最原始的 Cuscore 统计量，它假设初始的趋势是已知的。Q_{mm} 是使用局部平均估计值（EWMA）计算出的 Cuscore 统计量，Q_{mmlag} 是使用一个延迟的局部平均估计

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

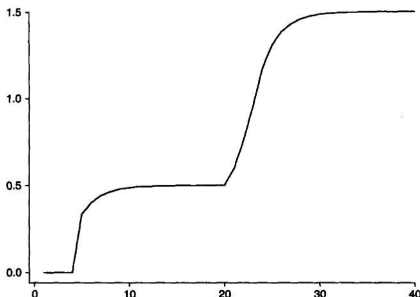


图 11-13 斜率系数估计值，$\beta_t = 0.25(EWMA_t - EWMA_{t-4})$

值计算得出的 Cuscore 统计量，$Qb1$ 是使用一个局部斜率估计值计算得出的 Cuscore 统计量，$Qb1_{lag}$ 是根据延迟的局部平均值，使用局部斜率估计值计算出的 Cuscore 统计量，而 Qb_{true} 则是使用了变化前后实际斜率系数计算出的 Cuscore 统计量。Cuscore 统计量的种类是很多的！

Q_{theo} 与 Qb_{true} 是理论上的基准值。我们希望一个有效的 Cuscore 统计量，与这两个基准值尽可能地接近。Qb_{true} 特别有趣。回过头看一看图 11-10。Qb_{true} 是将直线 AB 的偏差累积起来，因此当 $t=1$ 到 $t=20$ 来说，Qb_{true} 等于零。当 $t=21$ 的时候，我们将旧的斜率系数 $\beta=0.5$，转换为新的斜率系数 $\beta=1.5$，然后将直线 BC 与直线 AE 之间的偏差累积起来，其中直线 AE 是新的基准模型，同时假设斜率初始值是 $\beta=1.5$。

直线 AE 与直线 BC 平行（两条直线的斜率相同，$\beta=1.5$），因此当两者的偏差是常数时，Q 呈线性增长态势。在标准 Cuscore 统计量中，单个的偏差值会持续增加（不考虑噪声）。如果与标准的 Cuscore 统计量对比的话，Cuscore 统计量的合计数增长的速度比线性增长（Qb_{theo}）要快得多。Qb_{true} 比 Q_{theo} 在开始阶段更有优势，因为偏差在开始阶段就很大，识别变化的速度也

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

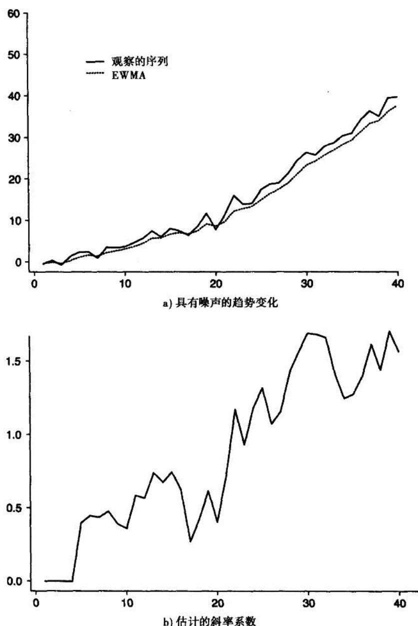


图 11-14 噪声影响下的斜率系数估计值，$\hat{\beta}_t = 0.25(EWMA_t - EWMA_{t-4})$

比较快。Qb_{theo} 的优势是两个斜率之间的相对大小和累积时间原点（也就是线段 AB 的长度）呈现函数关系。在我们识别突变过程中，变化发生之前持续

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

的时间越长，反馈到 Cuscore 统计量中的差异（$AE-BC$）就越大，这种方式超越标准 Cuscore 统计量的初始优势也就越大，因此能更快地识别出趋势的变化。在样本数据中，考虑噪声之后，产生了 $Qb1$ 和 Q_{mm} 两个 Cuscore 统计量。这两个统计量在实践中，哪个能准确的识别突变过程，取决于突变的动态变化和突变发生的周期。

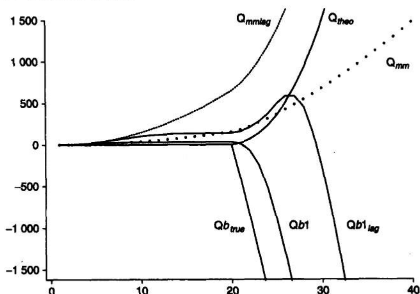


图 11-15 多个 Cuscore 统计量

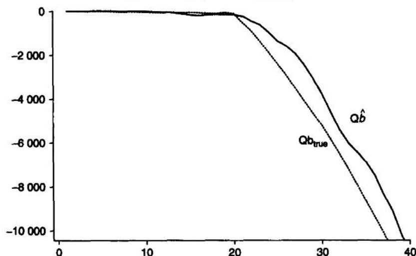


图 11-16 带有噪声数据的 Cuseore 统计量

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

之前我们认为 Q_{theo} 和 Qb_{true} 是理论上的基准，我们要得到的一个有效的 Cuscore 统计量，与这两个比较基准越接近越好。我们将“两个拖拉的方法相乘”得到的 Cuscore 统计量 $Q6$，与无噪声的数据非常接近。为什么会这样？在变化出现之后，观察值与局部平均估计值之间的差异（因为在持续不变的趋势中，EWMA 的表现是延迟的）乘上一个比较大的斜率估计值。这种 Cuscore 统计量的功效，对于某个特定类型的变化，不会造成拖拉的后果，却可以增强两者之间差异。这种方案可以应用在价格数据上吗？看完图 11-16 后，再做决定。此后，可以思考在识别的趋势是下降的状况，该怎么办。

在附录中的描述并不是非常正式。关于动态模型与变化识别的一些严谨的描述方法，可以参考波尔等人在 1994 年的著作。在那本参考书中，线性成长模型用参数表示局部平均值和趋势（成长），动态更新参数估计值与预测值，以及用正式的统计方法诊断参数（在这里就是斜率）的变化。对于股价数据来说，将动态线性模型的标准分布作为假设条件不是很恰当，主要是因为（假设呈现正态分布的）价格与回报呈现非常明显的非正态分布。虽然如此，不过如果我们采取一个规范的、“鲁棒性”非常好的方法，专注于均值估计，将标准差作为不确定性的一个参考值，并且完全不依赖正态分布（所谓的线性贝叶斯方法），这个模型是很有用的。

祝您在突变过程中捕获更多的猎物！

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

致　谢

格雷戈里·万·基普尼斯将我领进了统计套利的大门。在很大程度上，本书中的很多内容都是我们共同的成果，而且很高兴能报答他在智力上给予的帮助。我们的谈话经常从统计套利，偏离到宏观经济、通用科学，以及政治等议题。这些讨论并不都是从统计套利的角度进行考虑，可是我们有时候从一个不相关的议题中，得到一个在统计套利中有用的一个看法。现在要指出哪些观点是哪个人的，是一件不可能的事情。荣誉归功于万·基普尼斯，我对文中的内容负责。

John Wiley 出版公司的编辑和员工，尤其是 Bill Falloon，Emilie Herman，Laura Walsh，以及 Stacey Fischkelta，尽管我与他们没有见过面，但是在整个项目中，他们给予了很多帮助，在此对他们深表谢意。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

译 者 后 记

作为金融从业者，我们目睹了中国资本市场近10年来的快速发展，尤其是今年融资融券和股指期货的推出，使得做空机制从无到有，意义非凡。此时，能够有幸将《统计套利》一书译成中文，奉献给广大对此感兴趣的读者，我们颇感欣慰。

在整本书的翻译过程中，我们既有分工又有合作，从初稿到最终定稿，我们尽量做到字斟句酌，以呈现精品之心对待这部经典之作。陈雄兵翻译第1章、第6章、第7章、第8章、第9章、第10章、第11章。张海珊翻译前言、第2章、第3章、第4章、第5章。刘方圆对大部分章节进行了文字校对。为了更好地为广大读者服务，我们创办了统计套利网站（www.tongjitaoli.com），提供本书的勘误表以及与统计套利相关的资讯。读者也可以发送电子邮件给译者（邮箱：chenxiongbing@tsinghua.org.cn）进行联系。

译稿付梓，首先要感谢本书的作者安德鲁·波尔先生，正是他的作品使我们了解了什么是统计套利，了解了它的发展历程，看到了它对投资者可能发挥的重要作用。其次我们真诚地感谢机械工业出版社华章公司的颜诚若女士，感谢她采纳了这一选题并多次对译稿提出宝贵意见，使我们将这本书译成中文的愿望成真。最后，感谢中国农业银行执行董事、副行长潘功胜博士，能够在百忙之中拨冗为本书作序，感谢北京交通大学张秋生教授、中国农业银行张玉法先生、王哲先生，正是有了你们这些良师益友，才有了本书的顺利出版。

这是一本值得一读再读的著作。

译者

2010年10月

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

参 考 文 献

Arnold, V.I. *Catastrophe Theory*. New York: Springer-Verlag, 1986.

Bollen, N.P.B., T. Smith, and R.E. Whaley (2004). "Modeling the bid/ask spread: measuring the inventory-holding premium," *Journal of Financial Economics*, 72, 97–141.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Box, G.E.P., and G. Jenkins. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1976.

Box, G.E.P., and A. Luceno. *Statistical Control by Monitoring and Feedback Adjustment*. New York: John Wiley & Sons, 1987.

Carey, T.W. Speed Thrills. *Barrons*, 2004.

Engle, R. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1,008.

Fleming, I. *Goldfinger*. London: Jonathan Cape, 1959.

Gatev, E., W. Goetzmann, and K.G. Rouwenhorst. "Pairs Trading: Performance of a Relative Value Arbitrage Rule," *Working Paper 7032, NBER*, 1999.

Gould, S.J. *The Structure of Evolutionary Theory*. Cambridge: Harvard University Press, 2002.

Huff, D. *How to Lie With Statistics*. New York: W.W. Norton & Co., 1993.

Institutional Investor. "Wall Street South," *Institutional Investor*, March 2004.

Johnson, N.L., S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*, Volumes I and II. New York: John Wiley & Sons, 1994.

Lehman Brothers. *Algorithmic Trading*. New York: Lehman Brothers, 2004.

Mandelbrot, B.B. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. New York: Springer-Verlag, 1997.

Mandelbrot B.B., and R.L. Hudson. *The (Mis)Behavior of Markets: A Fractal View of Risk, Ruin, and Reward*. New York: Basic Books, 2004.

Orwell, George. *1984*. New York: New American Library, 1950.

Perold, A.F. (1988). "The Implementation Shortfall, Paper vs. Reality," *Journal of Portfolio Management*, 14:3, 4–9.

Pole, A., M. West, and J. Harrison. *Applied Bayesian Forecasting and Time Series Analysis*. New York: Chapman and Hall, 1994.

Poston, T., and I. Stewart. *Catastrophe Theory and its Applications*. London: Pitman, 1978.

Schack, J. (2004). "Battle of the Black Boxes," *Institutional Investor*, June 2004.

Sobel, D. *Longitude*. New York: Penguin Books, 1996.

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进:http://www.7help.net

统计套利

STATISTICAL ARBITRAGE

ALGORITHMIC TRADING INSIGHTS AND TECHNIQUES

在这本思路清晰、充满智慧、读起来又津津有味的著作中，作者以一种不凡而又有天分，却又不会让人难以亲近的方式，在兼顾学术的情况下，展现了一位经验丰富而成功的统计套利提倡者的见解。

——泽西岛尔米塔什资产管理有限公司数理研究与风险管理部主管 尼克·麦克劳德

多么惊人的发现啊！作者带给我们的是一个非同寻常的视角，让我们看到了通常被认为是无数对冲基金策略让人最难理解之处的历史和演化。他对统计套利基本驱动力的详细论述和睿智举例，对任何初次研究这一策略的人来说都是非常珍贵的资源，甚至我们这些旧时代的人也能从中受益匪浅。

——太平洋非主流资产管理公司执行董事 朱迪思·波斯尼科夫博士

作者以十分具有可读性和广泛性的方式展示了统计套利的历史。他根据自己十多年的从业经验，运用真实的案例，以编年史的方式记载了统计套利大行其道的年代，解释了统计套利最近为获利能力复兴所做的努力，并从艺术与科学的角度，为构建模型的新手提供了深入的视角。

——富兰克林街合伙公司投资组合经理 苏珊·卡德瑞贝克

《统计套利》一书以独特的视角带我们浏览了短期交易策略令人十分难以理解的世界。该书提供了一个完美的平衡，一边对短期技术交易策略的数学机理予以概念化，另一边则对这种策略近期的绩效从实践角度进行了较多的讨论。对于许多门外汉，甚至一些对冲基金专业人士来说，统计套利都是一个“黑匣子”策略。如果没有变“白”的话，作者也已经设法将“黑”变成了更亮的灰色光影。

——国际投资公司执行董事 克里斯汀·泰格森

作者具有广泛的统计套利交易的实践经验，同时也有能力非常清楚地解释统计套利的基础理论。因此，我将这本书隆重推荐给那些渴望掌握这一内容的读者。

——金融风险管理师 布鲁斯·洛克伍德




客服热线：
(010) 88379210, 88361066
购书热线：
(010) 68326294, 88379649, 68995259
投稿热线：
(010) 88379007
读者信箱：
hzjg@hzbook.com

华章网站 http://www.hzbook.com

网上购书：www.china-pub.com

WILEY
Publishers Since 1807
www.wiley.com




上架指导：金融投资
ISBN 978-7-111-32544-4
9 787111 325444
定价：38.00元